

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Instituto de Ciências Biológicas**  
**Programa de Pós-graduação em Microbiologia**

Thaís Inês Régis Moreira

**ANÁLISE DO CONJUNTO DE TRNA EM *NUCLEOCYTOVIRICOTA***

Belo Horizonte  
2023

Thaís Inês Régis Moreira

**ANÁLISE DO CONJUNTO DE TRNA EM *NUCLEOCYTOVIRICOTA***

Monografia de especialização apresentada ao Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Especialista em Microbiologia.

Orientador (a): Prof. Dr. Rodrigo Araújo Lima Rodrigues

Belo Horizonte  
2023

043

Moreira, Thaís Inês Régis.

Análise do conjunto de tRNA em Nucleocytoviricota [manuscrito] / Thaís Inês Régis Moreira. – 2023.

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

Orientador (a): Prof. Dr. Rodrigo Araújo Lima Rodrigues.

Monografia de especialização apresentada ao Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Especialista em Microbiologia.

1. Microbiologia. 2. RNA de Transferência. 3. Vírus Gigantes. I. Rodrigues, Rodrigo Araújo Lima. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 579



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
MICROBIOLOGIA

### ATA DE DEFESA DE MONOGRAFIA

Às 13:30 horas do dia 29 de setembro de 2023, reuniu-se, de forma remota a Banca Debatedora constituído pelo Dr. Mateus Sá Magalhães Serafim (Departamento de Microbiologia/ICB/UFMG) e o Prof. Rodrigo Araújo Lima Rodrigues (Departamento de Microbiologia/ICB/UFMG) - Orientador, para avaliar a Monografia intitulada "Análise do conjunto de Trnas em Nucleocyotovicota", da aluna Thaís Inês Régis Moreira Após a apresentação oral pública seguida de uma arguição, a aluna foi APROVADA, considerando as sugestões feitas pela Banca debatedora. Nada mais havendo a tratar, lavrou-se a presente ata, que será assinada pelos membros participantes da Banca Debatedora.

Belo Horizonte, 29 de setembro de 2023

Membros da banca:

Dr. Mateus Sá Magalhães Serafim  
Prof. Rodrigo Araújo Lima Rodrigues – Orientador

Profa. Nalu Teixeira de Aguiar Peres  
Coordenadora do Curso de Especialização em Diagnóstico Microbiológico



Documento assinado eletronicamente por **Rodrigo Araújo Lima Rodrigues, Professor do Magistério Superior**, em 05/10/2023, às 16:02, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Nalu Teixeira de Aguiar Peres, Professora do Magistério Superior**, em 05/10/2023, às 17:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mateus Sá Magalhães Serafim, Usuário Externo**, em 18/10/2023, às 11:22, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2692509** e o código CRC **29CA0D2E**.

## **AGRADECIMENTOS**

Agradeço primeiramente ao meu pai Carlos por ser tudo na minha vida. Não tenho palavras para descrever a quão grata eu sou por tê-lo como meu pai.

Ao meu companheiro Juliano e meus sogros Maria Lúcia e Deiber pelo incentivo, apoio e suporte na realização da especialização.

Ao meu orientador, Rodrigo, pela infinita paciência e disponibilidade para me ensinar, auxiliar e guiar na produção deste trabalho, e me incentivar na continuidade na micro.

Ao Programa de Pós-graduação em Microbiologia e a todos os professores da Especialização, que colaboraram com minha formação e aprendizado.

A minha pequena turma do curso de especialização pelas experiências, conhecimentos, finais de semana e descontrações compartilhadas que fez com que os dias fossem mais leves.

Por fim, agradeço a banca avaliadora pela disponibilidade em colaborar para a melhoria deste trabalho.

## RESUMO

Os vírus gigantes, tradicionalmente conhecidos como vírus grandes de DNA nucleocitoplasmáticos (NCLDV), fazem parte do novo filo estabelecido em 2023 pelo Comitê Internacional de Taxonomia de Vírus (ICTV), o filo *Nucleocytoviricota*. Os longos genomas dos vírus gigantes codificam genes envolvidos no processo de tradução de proteínas, como os RNA de transferência (tRNA) e as aminoacil tRNA sintetases (aaRS). Esses componentes não foram relatados anteriormente em outros vírus até o sequenciamento do genoma do primeiro vírus gigante de ameba, o Mimivirus bradfordmassiliense (originalmente denominado como *Acanthamoeba polyphaga mimivirus* - APMV). No presente trabalho foi possível fazer um estudo da prevalência dos tRNA a partir de uma pesquisa em larga escala buscando por genes de tRNA em genomas de grupos de vírus gigantes que infectam amebas de vida livre. Foram descobertos que das setenta sequências depositadas no GenBank até 31/12/2022 oriundas de vírus isolados, quarenta e oito continham ao menos um gene de tRNA. Essa distribuição se concentrou principalmente nos Tupanvírus, que são os detentores do maior e mais diverso aparato de tradução, e nos Mimivírus, grupo com um grande número de isolados. Encontrou-se grande diversidade na identidade do anticódon, mas foi observado uma tendência entre os vírus para aqueles correspondentes principalmente ao aminoácido leucina. A organização desses tRNA em agrupamentos também foi alvo de análise e ocorreu em vírus com elevada quantidade de tRNA codificados, como o Tupanvírus e o Yasminevírus. Os íntrons contidos nesses genes também foram avaliados e fizeram-se presentes em trinta e quatro genomas. Não foi possível traçar uma árvore filogenética dos tRNA em vírus gigantes de amebas, mas pode-se observar que, em sua maioria, as sequências apresentaram alta taxa de conservação entre os membros dos grupos ao qual pertencem. Os dados obtidos neste trabalho demonstram a grande diversidade e complexidade da maquinaria de tradução dos vírus gigantes. Análises envolvendo outros componentes do aparato traducional fornecerão uma visão geral dessa área ainda pouco explorada da genômica dos vírus gigantes.

Palavras-chave: aparato traducional; genes de tRNA; nucleocitovírus; vírus gigantes; tRNA.

## ABSTRACT

Giant viruses, commonly known as large nucleocytoplasmic DNA viruses (NCLDV), are part of the new phylum established in 2023 by the International Committee on Taxonomy of Viruses (ICTV), the phylum Nucleocytoviricota. The long genomes of giant viruses encode genes involved in the process of protein translation, such as transfer RNA (tRNA) and aminoacyl tRNA synthetases (aaRS). These components were not previously reported in other viruses until the sequencing of the genome of the first giant virus of amoeba, *Mimivirus bradfordmassiliense* (originally called *Acanthamoeba polyphaga Mimivirus* - APMV). In the present work, it was possible to study the prevalence of tRNA through a large-scale search for tRNA genes in the genomes groups of giant viruses that infect free-living amoebas. It was found that the seventy sequences deposited in GenBank from isolated viruses, forty-eight of them contained at least one tRNA gene. This distribution was highly present in the Tupanviruses, which have the largest translation apparatus, and the Mimiviruses, a group with a great number of isolates. A considerable diversity was found in the identification of the anticodon, however a tendency among the viruses towards those corresponding mainly to the amino acid leucine more evident. The organization of these tRNA into clusters was analyzed and occurred in viruses with a high number of encoded tRNA, such as Tupanvirus and Yasminevirus. The introns contained in these genes were also evaluated and were detected in thirty-four genomes. It was not possible to build a phylogenetic tree of the tRNA in giant viruses, despite being that, for the most part, the sequences showed a high rate of conservation among the members of the groups to which they belong. The data obtained in this study demonstrates the diversity and complexity of the translation machinery of giant viruses. Further analyses involving other components of the translational apparatus could provide a new overview of this still largely unexplored area of giant virus genomics.

Keywords: giant viruses; nucleocytovirus; tRNA genes; tRNA, translational apparatus.

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>8</b>
1.1 Os vírus e a dependência do hospedeiro .....	8
1.2 Vírus gigantes de DNA .....	9
1.3 Aparato traducional em vírus gigante .....	14
1.4 tRNA: biologia e função .....	16
<b>2. JUSTIFICATIVA</b> .....	<b>19</b>
<b>3. OBJETIVOS</b> .....	<b>20</b>
3.1 Objetivos geral.....	20
3.2 Objetivos específicos.....	20
<b>4. MATERIAIS E MÉTODOS</b> .....	<b>21</b>
4.1 Busca das sequências.....	21
4.2 Análises gerais dos genomas.....	21
4.3 Predição de tRNA.....	22
4.4 Análises gerais dos tRNA virais.....	22
4.5 Alinhamento múltiplo de sequências .....	23
<b>5. RESULTADOS</b> .....	<b>24</b>
5.1 Características gerais dos genomas analisados .....	24
5.2 Diversidade e caracterização de tRNA em vírus gigantes.....	25
5.3 Distribuição de tRNA virais .....	31
5.4 Localização dos tRNA nos genomas virais .....	33
5.5 Caracterização molecular de tRNA-leucina em vírus gigantes.....	36
<b>6. DISCUSSÃO</b> .....	<b>44</b>
<b>7. CONCLUSÃO</b> .....	<b>49</b>
<b>8. REFERÊNCIAS</b> .....	<b>50</b>
<b>9. ANEXOS</b> .....	<b>56</b>

# 1. INTRODUÇÃO

## 1.1 Os vírus e a dependência do hospedeiro

A microbiologia foi marcada pela descoberta de microrganismos, bactérias e fungos, como agentes causadores de doenças, ao final do século XIX. No campo da virologia o progresso foi um pouco mais tardio, muito justificado pela impossibilidade de visualização em microscópio óptico e cultivo *in vitro* dessa classe de agentes infecciosos (FRAENKEL-CONRAT, 1986; LECOQ, 2001). A descoberta do primeiro vírus, o Vírus do Mosaico do Tabaco (TMV), se iniciou em 1876 pelo químico agrícola Adolf Mayer, que demonstrou a natureza infecciosa do mosaico do tabaco, mas não conseguiu isolar e cultivar o patógeno, sugerindo então ser um tipo de bactéria. Em 1892, Dimitri Ivanovsky, publicou seu trabalho sobre suas experiências experimentais sobre esse vírus, destacando que mesmo após o extrato de folhas acometidas pela doença ser submetido a filtração, os sintomas ainda se manifestaram. Por tanto, sugeriu se tratar de uma toxina bacteriana, ou uma bactéria. Poucos anos depois, em 1898 e com as mesmas observações de Ivanovsky, Martinus Willem Beijerinck, relatou que o agente se multiplicava em associação com o metabolismo do hospedeiro e afirmou não se tratar de uma bactéria, denominando o novo agente infeccioso como “contagium vivum fluidum” (BOS, 1981, 1999; CREAGER, 2022; LECOQ, 2001). A partir dessas descobertas, o TMV seguiu sendo alvo de estudo durante o século XX, e a virologia ganhou um notável interesse dos pesquisadores, resultando em marcantes descobertas no campo.

Os vírus são ubíquos e as entidades mais abundantes na biosfera, com uma estimativa de  $10^{31}$  partículas existentes no planeta, que infectam formas de vidas em todos os reinos conhecidos (arqueias, bactérias, eucariotos, e até mesmo outros vírus) (BAR-ON; PHILLIPS; MILO, 2018; HENDRIX et al., 1999; RAOULT; FORTERRE, 2008). Sua principal característica é que são considerados parasitas intracelulares obrigatórios que, por não possuírem ribossomos, necessitam do recrutamento destas organelas do hospedeiro para traduzir seus RNA mensageiros (mRNA). Desta forma, dependem da célula hospedeira para obter as macromoléculas e a maquinaria de biossíntese necessária para replicação do seu genoma, produção de suas proteínas e montagem de suas partículas virais (MAYER et al., 2019; WALSH; MOHR, 2011). Embora as defesas imunes da célula hospedeira infectada por vírus inclui incapacitar

seu próprio aparato traducional, os vírus desenvolveram mecanismos capazes de evadir as respostas antivirais das células hospedeiras e de explorar suas funções metabólicas (MORENO-ALTAMIRANO; KOLSTOE; SÁNCHEZ-GARCÍA, 2019). Outras características comuns aos vírus é que eram tradicionalmente conhecidos por apresentarem pequenas dimensões (menor que 200 nm, de acordo com Andre Lwoff), e possuírem genomas pequenos, o que também significa poucos genes codificadores de proteínas (LOUTEN, 2016; LWOFF, 1957).

## 1.2 Vírus gigantes de DNA

As definições e características associadas aos vírus foram desafiadas pela descoberta do primeiro vírus gigante de DNA com uma extensão de genoma na gama das megabases, relatado por La Scola et al. (2003) (SCHULZ; ABERGEL; WOYKE, 2022). A circunstância de descoberta do *Mimivirus bradfordmassiliense* (originalmente denominado como *Acanthamoeba polyphaga mimivirus*), se iniciou em 1992, após um surto de pneumonia em Bradford, Inglaterra. Durante o estudo para investigar a fonte causadora do surto, foram isolados microrganismos de amostras de água, que por microscopia óptica e coloração de Gram, permitiram a visualização de células Gram-positivas de um organismo que chamaram de cocos de Bradford, dentro de uma espécie de ameba de vida livre, a *Acanthamoeba polyphaga* (Figura 1A) (CLAVERIE et al., 2009; ICTV, 2023; LA SCOLA et al., 2003; RAOULT et al., 2004; RODRIGUES et al., 2021).

Além da análise morfotintorial, foram realizadas outras técnicas como as moleculares e de isolamento e cultivo, mas todas falharam para a identificação do novo organismo (LA SCOLA et al., 2003). Durante 11 anos, os cocos de Bradford foram considerados bactérias intracelulares de amebas até que, por microscopia eletrônica de transmissão (TEM), foram visualizadas partículas icosaédricas de 400 nm de diâmetro recobertas por estruturas similares a longas fibrilas, fazendo com que o diâmetro total da partícula seja de cerca de 750 nm (Figura 1B). Também foi possível a observação da fase eclipse durante sua replicação, característica típica de ciclo viral (LA SCOLA et al., 2003; RAOULT et al., 2004; RODRIGUES et al., 2021). Além disso, o sequenciamento de seu genoma provou outra quebra de paradigma, exibindo um genoma de aproximadamente 1,181,404 pares de base e codificando mais de 1000 genes. A descoberta do Mimivírus, nome dado devido a sua capacidade de imitar

micróbios, mudou a percepção do tamanho e complexidade viral e até mesmo o conceito de vírus (CLAVERIE et al., 2009; RAOULT et al., 2004; RODRIGUES et al., 2021).

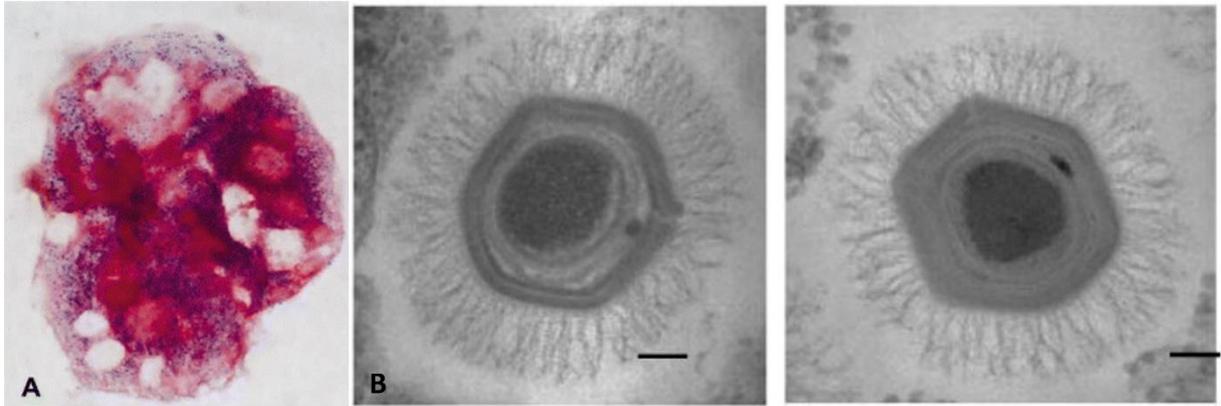


Figura 1. (A) Microrganismos (cocos de Bradford) em *Acanthamoeba polyphaga* em microscopia de luz após coloração de Gram. Fonte: Raoult et al., 2007 (B) *Mimivirus bradfordmassiliense* por microscopia eletrônica de transmissão. Fonte: Claverie et al., 2009

Devido a características morfológicas diferenciais das famílias virais já existentes, houve, em 2005, o estabelecimento da família *Mimiviridae* para inclusão dos Mimivírus (ICTV, 2005; LA SCOLA et al., 2005). Atualmente fazem parte da família os Tupanvírus, Mimivírus, Moumouvírus, Megavírus, Cotonvírus, Yasminevírus, Fadolivírus, Theiavírus e Rheavírus (ICTV, 2023; RODRIGUES et al., 2021). Os Mimivírus e a família *Mimiviridae* também foram incluídos ao grupo dos grandes vírus de DNA nucleocitoplasmáticos (NCLDVs), originalmente composto pelas famílias *Poxviridae*, *Phycodnaviridae*, *Iridoviridae* e *Asfarviridae* (IYER; ARAVIND; KOONIN, 2001; WILSON; VAN ETTEN; ALLEN, 2009). Nos anos seguintes a sua descoberta, outros vírus gigantes de amebas foram relatados, tiveram seus genomas sequenciados e foram incluídos ao expansivo grupo dos NCLDVs, como os Marseillevírus, Pandoravírus, Pithovírus, Cedratvírus, Orpheovírus, Pacmanvírus, Mollivírus, Kaumoebavírus, Medusavírus e Faustovírus (Figura 2) (KOONIN et al., 2020; KOONIN; YUTIN, 2019; MÖNTTINEN et al., 2021; RODRIGUES et al., 2021).

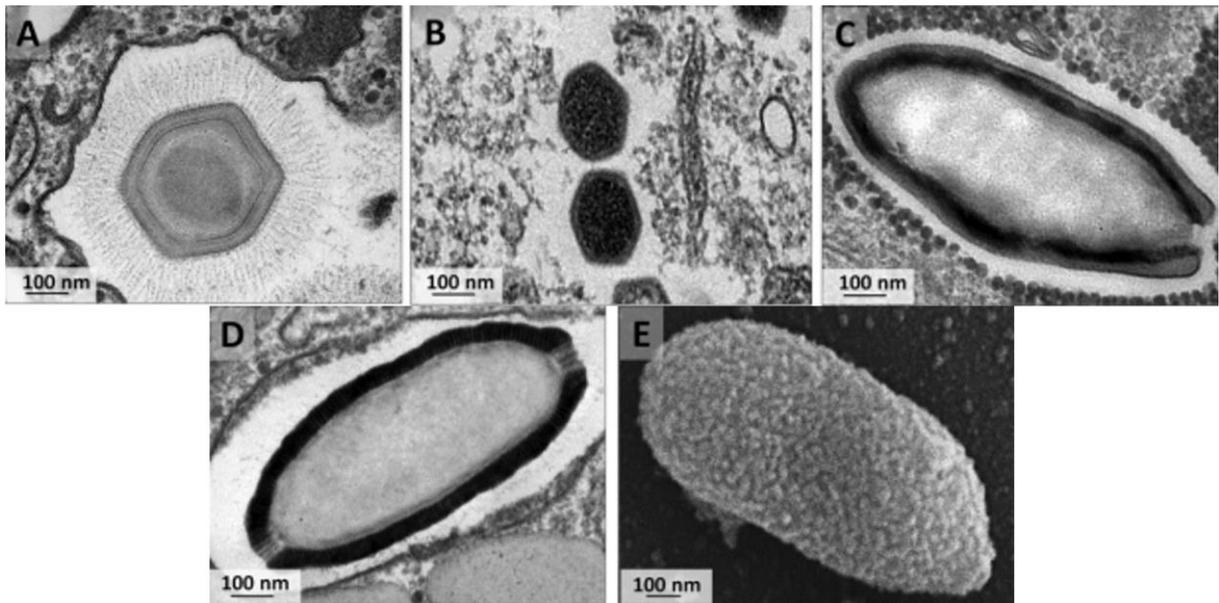


Figura 2. Imagens de microscopia eletrônica de transmissão e varredura de diferentes vírus gigantes de amebas do grupo dos grandes vírus de DNA nucleocitoplasmáticos (A) Mimivírus (B) Marillevírus (C) Pandoravírus (D) Cedratvírus (E) Orpheovírus. Fonte: Adaptado Rodrigues et al., 2021

Atualmente os NCLDVs foram oficialmente classificados em 2023 pelo Comitê Internacional de Taxonomia de Vírus (ICTV) (<https://ictv.global/taxonomy>) como um novo filo viral *Nucleocytoviricota* (NCVs), porém alguns, como os Pandoravírus, apesar de serem relacionados aos NCVs, permanecem sem classificação taxonômica. O filo *Nucleocytoviricota* é composto pelas classes *Pokkesviricetes* e *Megaviricetes*, anteriormente proposto por Colson et al. (2013) como *Megavirales*. A classe *Megaviricetes* abriga 3 ordens: *Algavirales* (composto pela família *Phycodnaviridae*), *Imitervirales* (composto pelas famílias *Allomimiviridae*, *Mesomimiviridae*, *Mimiviridae* e *Schizomimiviridae*) e *Pimascovirales* (composto pelas famílias *Iridoviridae*, *Ascoviridae*, *Marseilleviridae* e *Mamonoviridae*). Os *Pokkesviricetes* abrigam 2 ordens: *Asfuvirales* (composto pela família *Asfasviridae*) e *Chitovirales* (composto pela família *Poxviridae*) (Figura 3). A árvore filogenética concatenada do filo é baseada em 7 genes marcadores: DNA polimerase da família B (PolB), subunidade menor de RNA polimerase (RNAPS), enzima de empacotamento A32 (A32), topoisomerase da família II (TopoII), fator 3 de transcrição tardia de vírus (VLTF3), fator de transcrição IIB

(TFIIB) e uma helicase da superfamília II (SFII) (AYLWARD et al., 2021; ICTV, 2023; XIAN; XIAO, 2020).

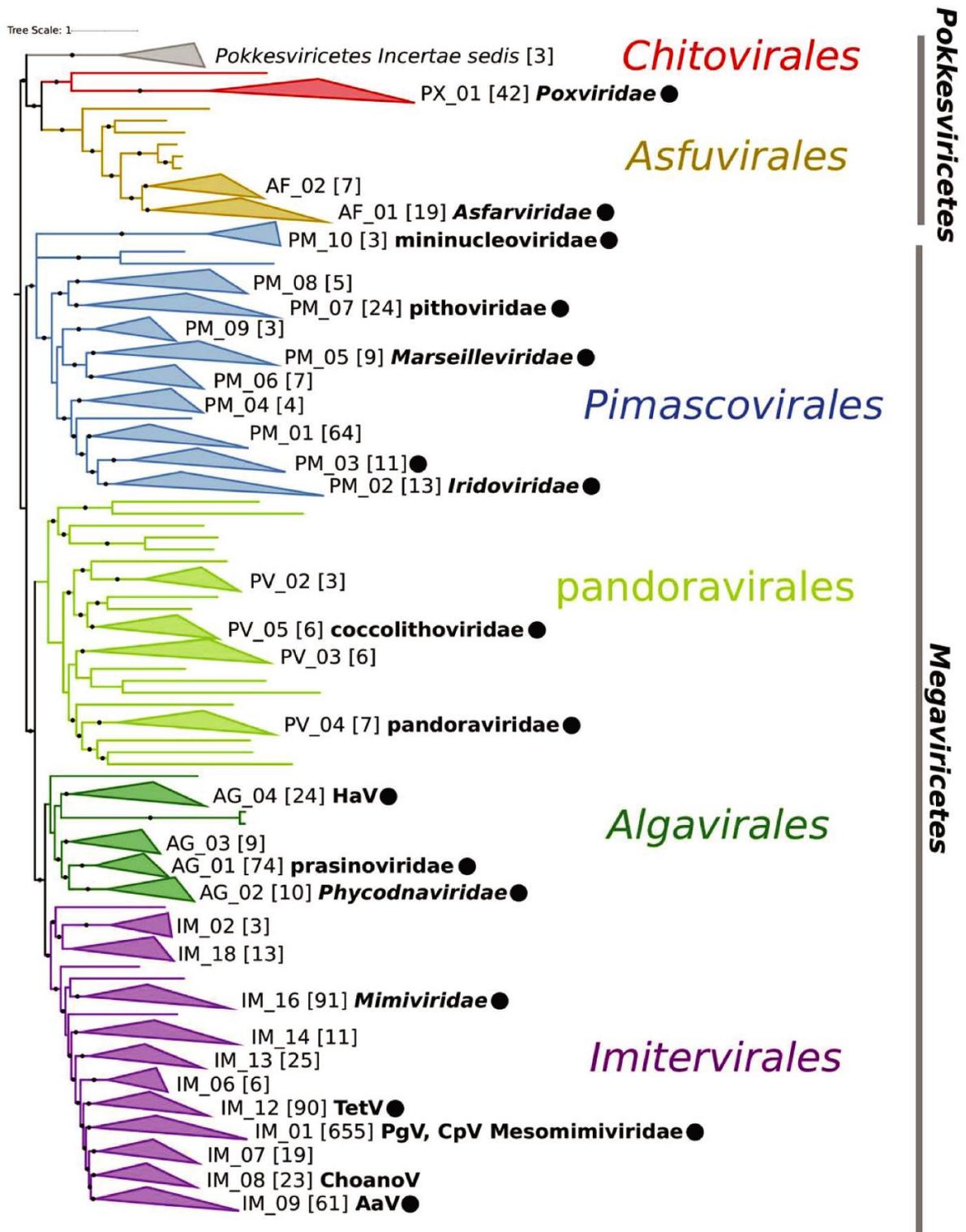


Figura 3. Árvore filogenética de *Nucleocytoviricota* com base no conjunto de 7 genes marcadores (DNA polimerase da família B (PolB), subunidade menor de RNA polimerase (RNAPS), enzima de empacotamento A32 (A32), topoisomerase da família

II (Topoll), fator 3 de transcrição tardia de vírus (VLTF3), fator de transcrição IIB (TFIIB) e uma helicase da superfamília II (SFII)). Círculos sólidos indicam grupos em que os vírus foram isolados, não somente identificados por metagenômica. As famílias são indicadas por clados recolhidos, com seu identificador não redundante fornecido à sua direita. O número de genomas em cada clado é fornecido entre colchetes. Os nomes de família estabelecidos são fornecidos em **negrito e itálico** e os nomes propostos são fornecidos em letras minúsculas. A presença de vírus cultivados notáveis é apresentada em **negrito** ao lado de alguns clados. Fonte: Aylward et al., 2021

Os nucleocitovírus compreendem um grupo de vírus de DNA fita dupla (dsDNA) detentores dos maiores genomas da virosfera, que variam em cerca de 100 quilo pares de base (kbp) a 2,5 mega pares de base (Mbp) e a maior quantidade de genes codificadores de proteínas relatados. Também são detentores das maiores partículas virais conhecidas, podendo chegar a medir 2,3  $\mu\text{m}$  de comprimento e 0,5  $\mu\text{m}$  de diâmetro, apresentam uma grande diversidade estrutural de suas partículas podendo exibir capsídeos icosaédricos, partículas esféricas, ovais ou similares a caudas (ABERGEL; LEGENDRE; CLAVERIE, 2015; ABRAHÃO et al., 2018). Os integrantes do filo infectam uma ampla gama de hospedeiros, de animais vertebrados e invertebrados a algas, amebas e outros eucariotos, e podem realizar todo o ciclo viral no citoplasma do hospedeiro, como é o caso dos Poxvírus, ou realizar a transcrição e replicação no núcleo, e a morfogênese no citoplasma, que devido a essa característica confere ao grupo o nome nucleocitovírus, como é observado para os Asfarvírus (Figura 4) (MÖNTTINEN et al., 2021; SCHULZ; ABERGEL; WOYKE, 2022).

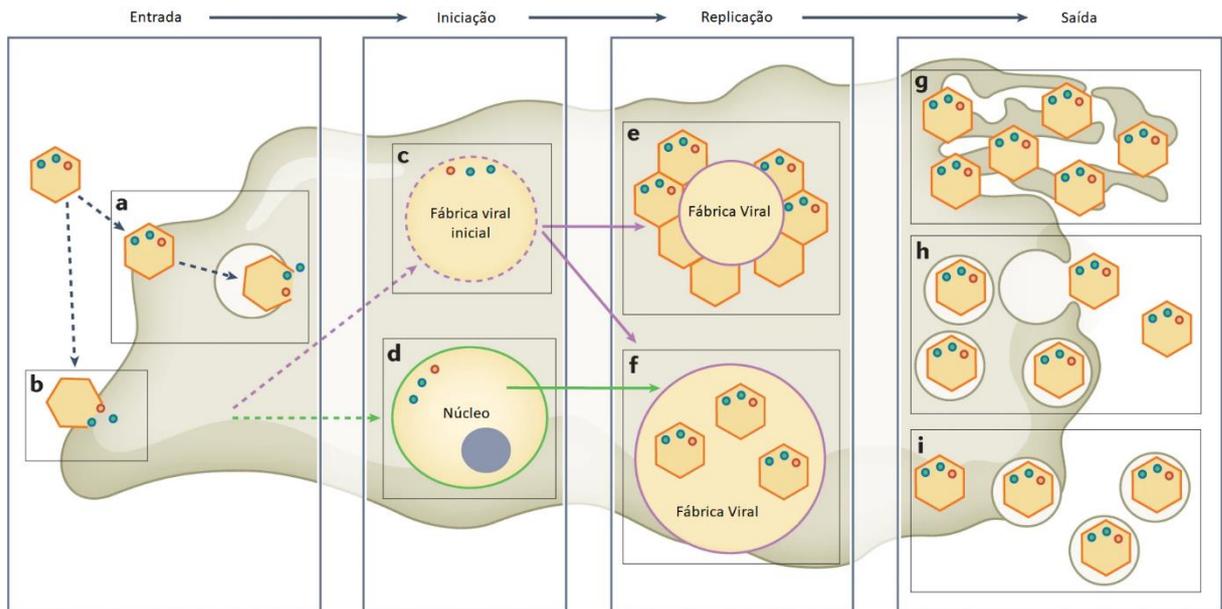


Figura 4. Ciclo viral de vírus gigantes com partículas >500nm (a) Endocitose (b) Fusão de membrana (c) transcrição no citoplasma (d) transcrição no núcleo do hospedeiro (e) morfogênese dos vírions na periferia da fábrica viral citoplasmática (f) morfogênese dos vírions na fábrica viral citoplasmática (g) liberação dos vírions por lise da célula hospedeira (h) exocitose (i) brotamento. Fonte: Adaptado de Schulz et al., 2022

A penetração dos vírus gigantes na célula hospedeira ocorre por fagocitose através do reconhecimento de partículas de ao menos 500 nm. Porém, para vírus que possuem um tamanho menor, como é o caso dos Marseillevírus que apresentam um diâmetro de aproximadamente 250 nm, eles foram capazes de desenvolver um mecanismo alternativo que, pela formação de vesículas gigantes podendo apresentar um tamanho de até mais de 1  $\mu\text{m}$  de diâmetro, as torna apropriadas para a fagocitose. Além disso, foi descrito que os Marseillevírus também exploram outras vias endocíticas para penetração de suas partículas individuais (ARANTES et al., 2016; RODRIGUES et al., 2021; SAHMI-BOUNSIAR et al., 2021; WEISMAN; KORN, 1967).

### 1.3 Aparato traducional em vírus gigante

Os nucleocitovírus, diferentemente dos vírus menores que dependem do maquinário da célula hospedeira, exibem certa independência sobre seus hospedeiros, uma vez que possuem genes relacionados a proteínas necessárias ao processo de transcrição, como RNA polimerases, helicases e enzimas capeadoras de

mRNA, e ao processo de tradução dos seus genomas, como os RNA de transferência (tRNA), as aminoacil tRNA sintetases (aaRS) e outros fatores de tradução (FT) (KOONIN; YUTIN, 2010; RAOULT et al., 2004).

Dessa forma, os Mimivírus foram os primeiros vírus a serem relatados como detentores de genes para componentes envolvidos no processo de tradução. O primeiro isolado foi descrito com uma presença de genes codificadores de seis tRNA, quatro tipos diferentes de aaRS, três fatores de iniciação e alongamento, e um fator de liberação da cadeia peptídica (RAOULT et al., 2004). Com o avanço dos estudos acerca do genoma viral e dos tRNA e aaRS dos NCVs, outros isolados da família *Mimiviridae* foram descritos exibindo genomas com um número cada vez maior de componentes do aparato traducional. Os Klosneuvírus foram descritos por Schulz et al. (2017) com um amplo número de componentes do sistema de tradução com a presença de 25 tRNA e 19 aaRS, além de 11 fatores de iniciação e alongamento da tradução, um fator de liberação da cadeia peptídica e várias enzimas modificadoras de tRNA, superando o *Mimivirus bradfordmassiliense* tanto em quantidade de componentes, como em diversidade destes.

Porém, atualmente a família *Mimiviridae* conta com os Tupanvírus, isolados e descritos por Abrahão et al. (2018), e o mais recente Yasminevírus, isolado e descrito por Bajrai et al. (2020) como os maiores detentores da virosfera de componentes de aparato traducional, ambos com até 70 tRNA e 20 aaRS, 11 fatores para todas as etapas de tradução e fatores relacionados à maturação de tRNA e mRNA, bem como associados a modificação de proteínas ribossomais. De todos os componentes do aparato de tradução, ainda não foi relatado em nenhum vírus a presença de genes para os ribossomos.

Estudos indicam que a presença de genes relacionados à replicação, transcrição e/ou tradução nesses vírus, conferiria aos vírus possibilidades de se adaptarem frente ao viés de uso de códons (CUB) do hospedeiro que devido a fatores, como por exemplo composição e conteúdo de GC do genoma, posição e contexto dos códons nos genes, dobramento do mRNA, abundância e interações do tRNA dos genomas entre o hospedeiro e o vírus infectante, fazem com que esse viés seja diferente e haja a necessidade do vírus de favorecer ao seu próprio CUB com o objetivo de otimizar a síntese de proteínas e/ou ampliar a gama de hospedeiros possíveis (DUNCAN; DUNIGAN; VAN ETEN, 2020; MORGADO; VICENTE, 2019b; PARVATHY; UDAYASURIYAN; BHADANA, 2022).

## 1.4 tRNA: biologia e função

O RNA de transferência (tRNA) é uma cadeia curta de nucleotídeos que possuem uma faixa de comprimento entre 70 e 100 nucleotídeos (nt). Os tRNA em sua forma secundária são estruturalmente comparados ao formato de trevo composto por uma haste aceptora, um braço D, um braço T, um braço variável e um braço do anticódon (Figura 5A). Cada braço consiste em uma haste de fita dupla e um loop de fita simples, exibindo um padrão de haste-loop. Há uma alta conservação nas estruturas da haste aceptora, do braço do anticódon e do braço T, ao passo que o braço D e o braço variável podem variar quanto ao seu tamanho. Normalmente é encontrado na haste aceptora 7 nt, no braço T e no braço do anticódon 5 nt, no braço D entre 3 e 5 nt, e no braço variável entre 3 e 21 nt, sendo a porção menos conservada no tRNA. A variação no braço variável promoveu a classificação dos tRNA em duas classes, sendo a classe I caracterizada por braços curtos, de 4 ou 5 nucleotídeos presentes, enquanto a classe II apresenta regiões longas, a partir de 10 nucleotídeos presentes. Sua estrutura terciária é em formato de L devido ao empilhamento da haste aceptora com o braço T, dobradas por uma hélice dupla contínua, e do braço D com o braço do anticódon (Figura 5B). As interações entre o loop D e o T garantem a estabilidade da estrutura geral da molécula (KRAHN; FISCHER; SÖLL, 2020; ROBERT W. HOLLEY et al., 1965; SIGLER, 1975)

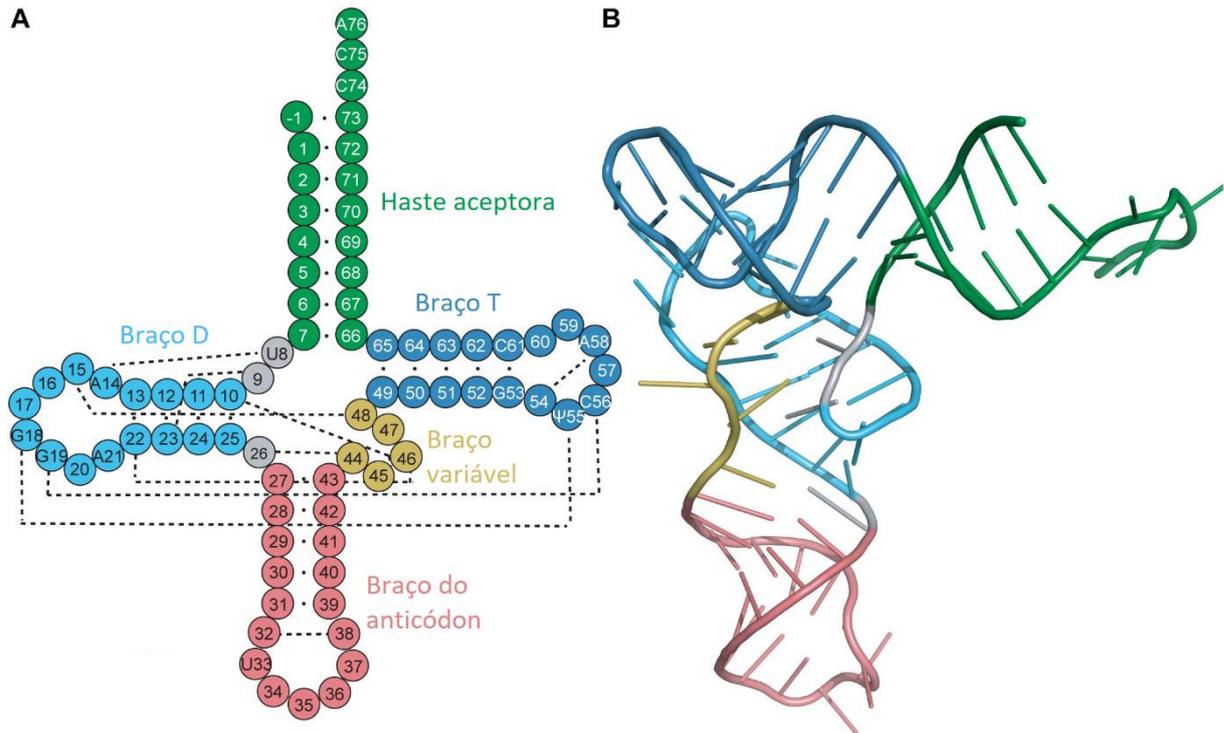


Figura 5. Estrutura do tRNA (A) Representação secundária com a numeração padrão de tRNA cuja haste aceptora é representada em verde, o braço T em azul escuro, o braço D em azul claro, o braço variável em amarelo, o braço anticódon em rosa. Linhas tracejadas correspondem às interações terciárias; (B) representação terciária de uma molécula de tRNA. Fonte: Adaptado Krahn et al.,2020

Os tRNA são moléculas adaptadoras que exercem um papel central na biossíntese de proteínas, fazendo a decodificação do código genético dos mRNA em aminoácidos. As interações de uma das extremidades L da estrutura terciária do tRNA, o anticódon, com o códon correspondente do mRNA, determinam a sequência de aminoácidos na cadeia polipeptídica. A outra extremidade, correspondente a haste aceptora e o braço D, fazem o reconhecimento das enzimas aminoacil tRNA sintetase que, por sua vez, fazem reações de aminoacilação na haste aceptora para executarem a ligação covalente dos aminoácidos corretos à extremidade 3' dos seus tRNA cognatos. O braço T auxilia em uma eficiente tradução a partir das interações entre os tRNA e os ribossomos (LYONS; FAY; IVANOV, 2018; RAINA; IBBA, 2014; SCHIMMEL, 2017).

Além da atuação na tradução de proteínas, que é o papel canônico dos genes tRNA, novos papéis e funções emergentes foram relatadas para os tRNA em diferentes processos metabólicos e celulares como na síntese da parede celular em

bactérias (STEWART; ROBERTS; STROMINGER, 1971), iniciadores de transcriptase reversa em vírus (MARQUET et al., 1995), precursor na biossíntese de heme e clorofila em plantas (JAHN; VERKAMP; SÖLL, 1992), e na biossíntese de aminoácidos dependente de tRNA (SHEPPARD et al., 2008). Especificamente nos vírus, os tRNA atuam como cofatores metabólicos e iniciadores na regulação da tradução, empacotamento e iniciação da transcrição reversa, e também participam de diversas funções biológicas na infecção (ALBERS; CZECH, 2016; MORGADO; VICENTE, 2019b; OLIVEIRA et al., 2022). Os tRNA são encontrados em todos os organismos celulares em abundância.

Curiosamente, alguns grupos de vírus também apresentam genes que codificam para diferentes tRNA, como os baculovírus, alguns bacteriófagos e virófagos (DELESALLE et al., 2016; OLIVEIRA et al., 2022; RODRIGUES et al., 2021). Contudo, foi com a descoberta dos vírus gigantes que o arsenal de tRNA na virosfera ganhou destaque diante da grande diversidade e abrangência destas moléculas nestes vírus.

## 2. JUSTIFICATIVA

Durante muito tempo os vírus foram considerados entidades simples, carregando apenas um ácido nucléico (DNA ou RNA) que não codifica um aparato de tradução de proteínas e não sintetizam ATP, o que os torna parasitas intracelulares obrigatórios completamente dependentes dos seus hospedeiros para sua multiplicação, liberação de sua progênie e disseminação. Porém, com a descoberta dos vírus gigantes, essas características mudaram, e foi observado que muitos de seus genes codificam proteínas e componentes envolvidos no processo de tradução não observados anteriormente em outros vírus e até então considerados exclusivos de organismos celulares, como por exemplo, RNA transportadores e aminoacil-tRNA-sintetases. Com o isolamento dos Tupanvírus, o número de genes observados até então aumentou e juntamente com a descrição da expressão de diferentes características nunca antes vistas em outros vírus, revelou ser o vírus gigante com o aparato traducional mais completo da virosfera. Mais recentemente em 2020, ocorreu o isolamento do Yasminevírus que demonstrou a elevada presença de genes envolvidos na tradução.

O avanço nas áreas de genômica e transcriptômica e os crescentes estudos aplicados aos vírus gigantes é de fundamental importância para caracterização de novos isolados e fundamentação para questões ainda em discussão na virologia, como a função da presença desses genes, bem como sua origem e evolução. Em especial, os tRNA de vírus gigantes não são amplamente estudados e explorados, portanto existe uma carência na literatura que descreva e compare os dados dos vírus gigantes que infectam amebas de vida livre acerca dos seus tRNA bem como a relação que esse componente pode fornecer na compreensão da evolução dos nucleocitovírus. Portanto, este conjunto de análises tem como objetivo fazer uma compilação de dados de genômica e transcriptômica de vírus gigantes de amebas de modo a estabelecer e identificar de forma mais clara as diferenças e similaridades de fatores comuns entres esses vírus contribuindo assim para facilitar o avanço de futuros trabalhos.

### **3. OBJETIVOS**

#### **3.1 Objetivos geral**

Avaliar a prevalência e diversidade dos tRNA nos grupos dos vírus gigantes que infectam amebas de vida livre.

#### **3.2 Objetivos específicos**

- Analisar e comparar as características genômicas de nucleocitovírus;
- Fazer a predição de genes tRNA de nucleocitovírus;
- Analisar e comparar as características de tRNA, e avaliar sua distribuição;
- Avaliar a presença de íntrons nos genes de tRNA e sua distribuição, e analisar as características dessas sequências;
- Avaliar a conservação dos genes de tRNA entre os vírus dos grupos virais.

## **4. MATERIAIS E MÉTODOS**

### **4.1 Busca das sequências**

Como métodos de busca foram pesquisados os depósitos com genomas completos até a data de 31/12/2022 do grupo dos vírus gigantes que infectam amebas de vida livre e que tenham sido devidamente isolados. Os genomas de diferentes isolados de Yasminevírus (BAJRAI et al., 2020), Tupanvírus (ABRAHÃO et al., 2018), Mimivírus (RAOULT et al., 2004), Kaumoebavírus (BAJRAI et al., 2016), Pacmanvírus (ANDREANI et al., 2017), Faustovírus (RETENO et al., 2015), Marseillevírus (BOYER et al., 2009), Orpheovírus (ANDREANI et al., 2018), Cedratvírus (ANDREANI et al., 2016), Pithovírus (LEGENDRE et al., 2014), Pandoravírus (PHILIPPE et al., 2013), Mollivírus (LEGENDRE et al., 2015), Medusavírus (YOSHIKAWA et al., 2019) e Clandestinovírus (ROLLAND et al., 2021), foram recuperados em formato fasta do GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) (Anexo 1). Para o Clandestinovírus e o Yasminevirus saudimassiliense foi constatado que os genomas completos de cada um estavam divididos em dois depósitos, portanto ambos foram considerados para as análises posteriores. O acesso aos genomas que fazem parte do patrimônio genético brasileiro está devidamente registrado no SISGEN sob o código de acesso A8952E4.

### **4.2 Análises gerais dos genomas**

Os dados das características do genoma, como taxonomia, tamanho e o número de acesso foram obtidos do GenBank. Dados como porcentagem de conteúdo GC do genoma foram obtidos do software ARAGORN v.1.2.41 (LASLETT; CANBACK, 2004). Com exceção do Yasminevirus saudimassiliense e Clandestinovirus cujos genomas estavam depositados em dois códigos de acessos diferentes, em que foram compilados e considerados o tamanho e o conteúdo de GC do genoma de acordo com o descrito por seus autores originais Bajrai et al. (2020) e Rolland et al. (2021), respectivamente. O conjunto de dados e informações dos genomas foram anotados em planilhas usando o Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA), dispostos e organizados por afiliação filogenética dos grupos de vírus. O conjunto de dados genômicos foram plotados e analisados utilizando o GraphPad Prism 8.0 (GraphPad Software Inc., San Diego, CA)

### 4.3 Predição de tRNA

As sequências dos genomas recuperados foram submetidas individualmente ao software ARAGORN v.1.2.41 no formato FASTA utilizando os seguintes parâmetros: tipo tRNA; permitir íntrons; sequência linear; fita simples e dupla (LASLETT; CANBACK, 2004). Como critério de inclusão para análises comparativas posteriores, foram selecionados aqueles genomas cujo ARAGORN retornou ao menos 1 tRNA. Em alguns casos o isotipo e o anticódon previsto por ARAGORN não foram discriminados com precisão, isto é, o software indicou dois isotipos para um mesmo gene tRNA, sendo denominados como tRNA ambíguos. Nesses casos, foi feita uma reanálise utilizando outro programa de predição de tRNA, o tRNA-scanSE 2.0 (CHAN; LOWE, 2019). Sendo assim, genomas que não apresentaram genes de tRNA, ou apresentaram tRNA ambíguos e/ou indeterminado foram excluídos do estudo. Os dados de porcentagem de conteúdo GC, presença e tamanho de íntrons, foram obtidos pela predição do software ARAGORN. Os dados e informações dos genes de tRNA virais foram anotados em planilhas usando o Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA) e o conjunto de dados foram plotados e analisados utilizando o GraphPad Prism 8.0 (GraphPad Software Inc., San Diego, CA)

### 4.4 Análises gerais dos tRNA virais

Os dados acerca dos genes de tRNA virais como porcentagem de GC, presença e tamanho de íntrons, foram obtidos pelo software ARAGORN (LASLETT; CANBACK, 2004). Os dados e informações dos genes de tRNA virais também foram anotados em planilhas usando o Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA). O conjunto de dados foram plotados e analisados utilizando o GraphPad Prism 8.0 (GraphPad Software Inc., San Diego, CA)

Para análise da posição do gene no genoma e a fita em que é expresso, bem como de agrupamentos de genes de tRNA foram selecionados um representante de cada grupo viral que atendesse a condição de tRNA > 1, com exceção do gênero *Mimivirus*, do qual foi selecionado um representante para cada linhagem (*Mimivirus*, *Moumouvirus*, *Megavirus*). Os genes de tRNA foram considerados agrupados se apresentassem densidade genética de tRNA > 1 em até 500bp. O servidor PROKSEE

(<https://proksee.ca/projects>) foi utilizado para comparação e representação gráfica dos dados (STOTHARD; GRANT; VAN DOMSELAAR, 2019).

#### **4.5 Alinhamento múltiplo de sequências**

As sequências de genes de tRNA dos vírus gigantes identificadas pelo software ARAGORN foram alinhadas utilizando o software Muscle com parâmetros padrão, implementado no programa MEGA 11 (EDGAR, 2004; TAMURA; STECHER; KUMAR, 2021). Genes com a presença de íntrons tiveram essas regiões removidas a partir da sequência predita pelo ARAGORN. Foi utilizado o gene de tRNA-leucina para exibir o alinhamento entre as sequências de nucleotídeos e selecionado um representante de cada grupo viral que atendesse a condição de tRNA-leucina  $\neq 0$ .

## 5. RESULTADOS

### 5.1 Características gerais dos genomas analisados

Foram recuperados do GenBank um total de setenta genomas completos (Anexo 1). O grupo que apresentou maior prevalência de genomas nesse estudo foram os Mimivírus com 21 genomas (30%) seguido pelos Faustovírus com 17 genomas (24,2%), Pandoravírus e Marseillevírus com 8 genomas cada (11,4%), Kaumoebavírus, Cedratvírus, Mollivírus, Pacmanvírus, Tupanvírus e Medusavírus apresentaram 2 genomas (2,9%), e o Clandestinovírus, Orpheovírus, Yasminevírus e Pithovírus apresentaram somente 1 genoma cada (1,4%) (Figura 6A).

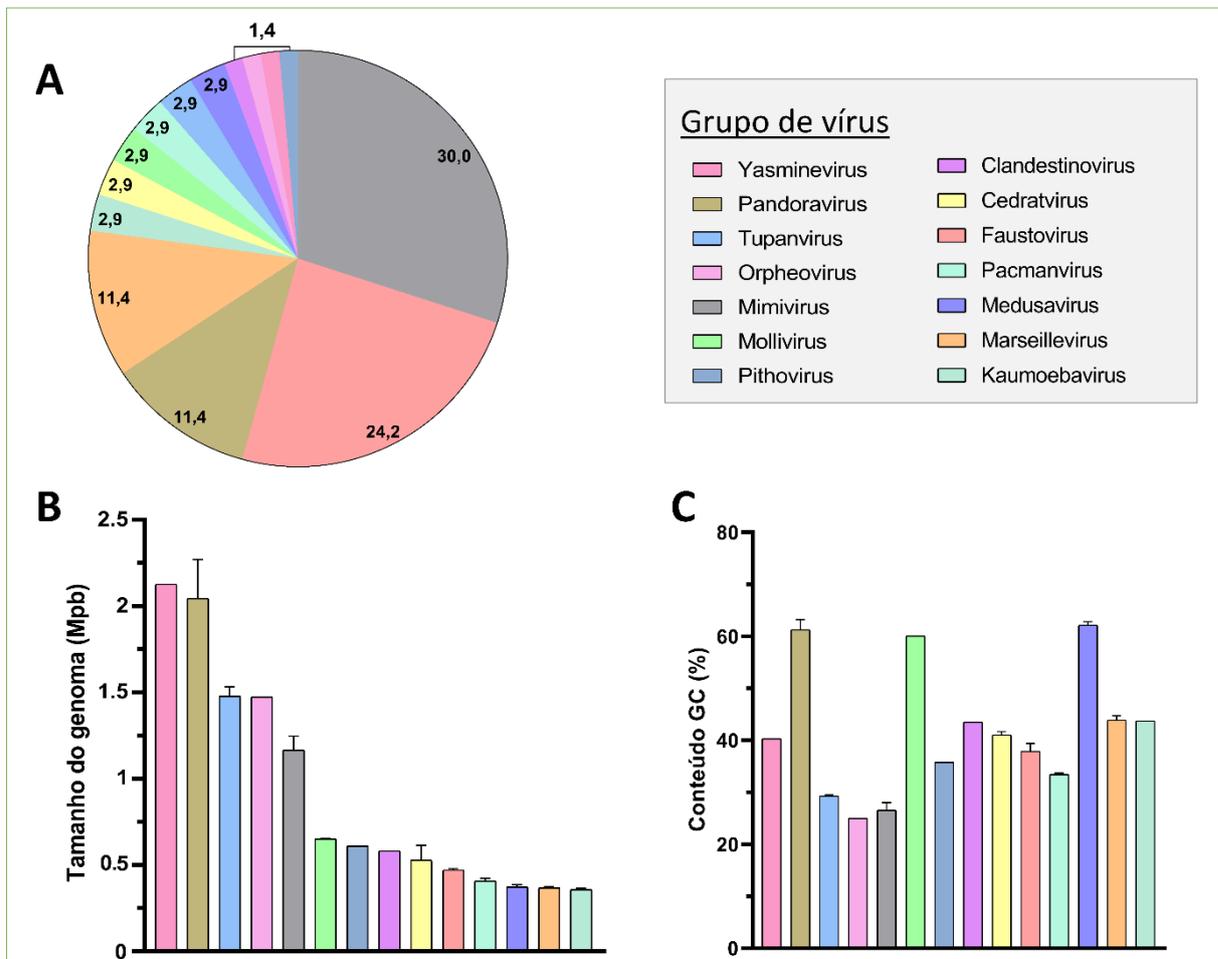


Figura 6. Características dos genomas dos grupos virais. (A) Distribuição de genomas nos grupos virais; (B) Média do tamanho dos genomas por grupo viral; (C) Média do conteúdo GC dos genomas por grupo viral.

O tamanho destes genomas variou de 350,731 kbp (Kaumoebavirus Sc), a 2.473,870 kbp (Pandoravirus salinus), e o conteúdo GC desses genomas variou de 24,6% (Moumouvirus moumou) a 63,7% (Pandoravirus dulcis). Dentre os grupos de vírus analisados, o Yasminevírus apresentou a maior média quanto ao tamanho do genoma com um valor igual a 2.120 kbp (Figura 6B) e um conteúdo GC de 40,3% (Figura 6C), seguido dos Pandoravírus com 2.040 kbp sendo 61,3% de conteúdo GC. Além disso, os Pandoravírus apresentaram um alto desvio padrão de um conjunto de genomas que variaram de 2.474 kbp a 1.787 kbp, resultando em um desvio padrão de  $\pm 224,892$  kbp. O Tupanvírus e o Orpheovírus apresentaram uma média muito próxima (1.477,887 kbp e 1.473,573 kbp, respectivamente) com uma diferença de apenas 4,314 kbp entre eles. Ambos também apresentaram uma das menores médias de conteúdo GC, sendo o Tupanvírus com 29,5%, e o Orpheovírus com o menor valor entre os grupos de 25% que, com apenas um isolado incluído nesse estudo, está condizente com o relatado previamente (ANDREANI et al., 2018).

Além disso, outros grupos que apresentaram médias adjacentes de tamanho de genomas foram os Medusavírus (372 kbp) e Marseillevírus (368 kbp), com uma diferença de 4,495 kbp. Tal semelhança não se manteve para o conteúdo GC, onde os Medusavírus apresentaram o maior resultado dentre todos os grupos, igual a 62,2%, enquanto os Marseillevírus exibiram 43,9% de GC. Os Mimivírus revelaram uma média de genoma de 1.160 kbp e conteúdo GC de 26,6%, próximo aos Tupanvírus e Orpheovírus. Os Mollivírus, apesar do baixo tamanho do genoma,  $\pm 650$  kbp, e um baixo valor de desvio padrão nessa análise, exibiram um alto teor de conteúdo GC, igual a 60,1% com nenhum desvio padrão, indicando que o grupo manteve o mesmo valor de %GC entre eles. O menor genoma mostrou ser pertencente aos Kaumoebavírus com aproximadamente 357 kbp e um conteúdo GC de 43,7%.

## 5.2 Diversidade e caracterização de tRNA em vírus gigantes

Dentre os setenta genomas recuperados, 48 retornaram ao menos 1 tRNA. Portanto, os demais foram excluídos das análises subsequentes. O genoma das espécies Tupanvirus salinum, Yasminevirus saudimassiliense e Medusavirus medusae apresentaram genes de tRNA cujo ARAGORN não conseguiu fazer a

predição de forma fidedigna, isto é, gerou predições ambíguas, portanto não foram incluídos no estudo.

Portanto, os 48 genomas codificam um total de 464 tRNA (Figura 7A), apresentando uma diversidade de 61 dos 64 anticódons para os 22 diferentes aminoácidos (SRINIVASAN; JAMES; KRZYCKI, 2002) e 1 tRNA de parada (TTA). Entre os grupos virais, a quantidade total de tRNA variou de 1 a 138. Os Tupanvírus apresentaram a maior quantidade dentre os grupos do estudo somando um total de 138 tRNA (29,7%) sendo 71 tRNA em *Tupanvirus salinum* e 67 em *Tupanvirus altamarinese*. O segundo grupo foi o dos Mimivírus com 126 (27,2%), variando de 9 (*Moumouvirus australiense* e *Megavirus baoshanense*) a 3 tRNA (*Moumouvirus moumou*). A maior incidência foi de 7 tRNA, presente em 8 dos 21 Mimivirus (*Mimivirus terra2*, *Acanthamoeba castellanii mamavirus*, *Hirudovirus*, *Samba vírus*, *Mimivirus Bombay*, *Mimivirus bradfordmassiliense*, *Acanthamoeba polyphaga mimivirus strain Oyster* e *Mimivirus reunion Queen*) e 5 tRNA, presente em 7 isolados (*Megavirus powaiense*, *Megavirus vitis*, *Megavirus Iba*, *Mimivirus U306*, *Mimivirus lagoaense*, *Moumouvirus maliensis* e *Moumouvirus goulettemassiliense*). Ainda foram visualizados 4 tRNA em 2 vírus (*Megavirus chilense* e *Acanthamoeba polyphaga mimivirus isolate M4*) e apenas 1 (*Megavirus courdo11*) com 6 tRNA.

Ademais, os Pandoravírus apresentaram 97 tRNA (20,9%) sendo 4 vírus (*Pandoravirus inopinatum*, *Pandoravirus celtis*, *Pandoravirus macleodensis* e *Pandoravirus neocaledonia*) com 9 tRNA cada e os demais variaram em 11 (*Pandoravirus quercus*), 15 (*Pandoravirus japonicus DNA*), 17 (*Pandoravirus salinus*) e 18 tRNA (*Pandoravirus dulcis*). *Yasminevírus* exibiu 70 tRNA (15,1%) para o único isolado do grupo, *Yasminevirus saudimassiliense*. *Mollivírus* exibiu 11 tRNA totais no grupo (2,4%, sendo 5 para o *Mollivirus sibericum* e 6 para o *Mollivirus kamchatka*). *Clandestinovírus* exibiu apenas 1 tRNA (0,22%), antecedido pelo *Pacmanvírus*, *Orpheovírus*, *Cedratvírus* com 2 tRNA (0,43%) cada, e o *Faustovírus* com 3 tRNA (0,65%). Estes, por apresentarem um baixo valor que acarretou uma porcentagem menor que 1, foram representados como outros (2,4%) para preservar uma melhor visualização do gráfico (Figura 7A).

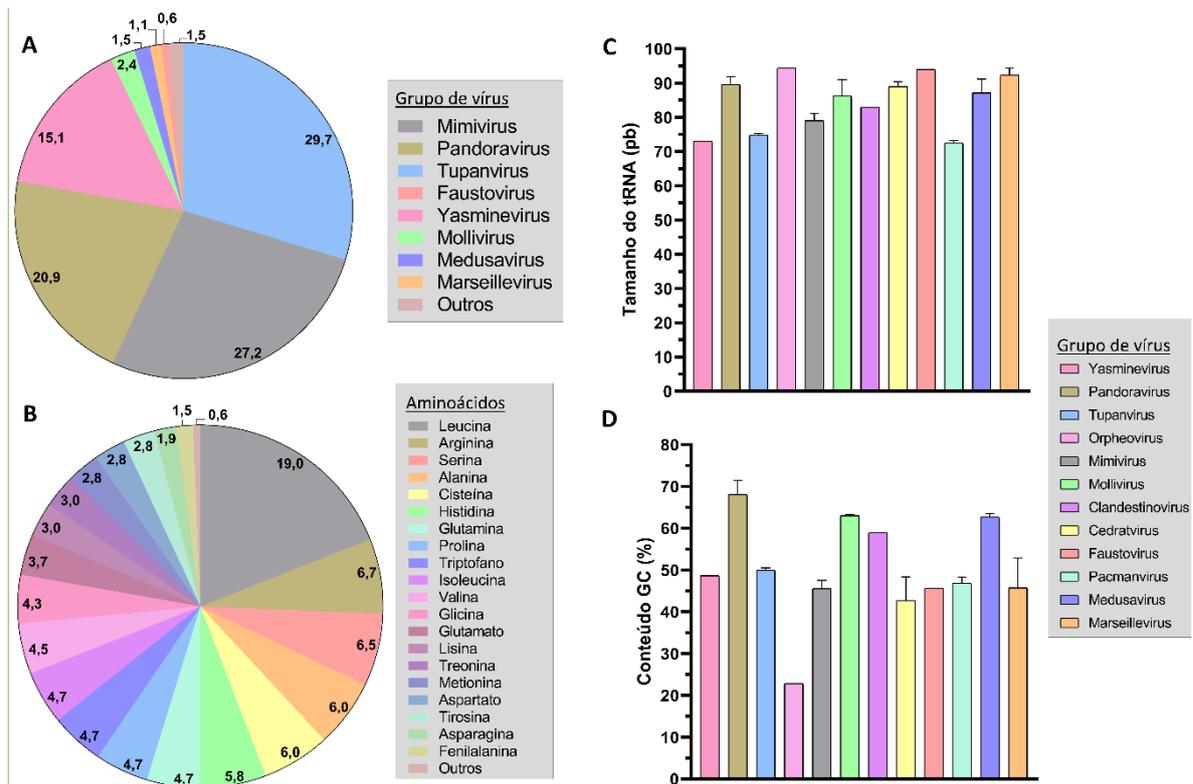


Figura 7. Análise de tRNA em vírus gigantes. (A) Distribuição de tRNA nos grupos virais; (B) Prevalência dos aminoácidos no estudo; (C) Média do tamanho dos tRNA por grupo viral; (D) Média do conteúdo GC dos tRNA por grupo viral.

Em seguida, a quantidade de tRNA considerando o aminoácido cognato variou de 88 a 1, sendo para a leucina a de maior prevalência (19%), e os tRNA não canônicos, selenocisteína e pirrolisina (KRAHN; FISCHER; SÖLL, 2020), o de menor, com apenas 1 tRNA (0,2%) cada, juntamente com o tRNA de parada. Estes genes foram encontrados respectivamente em Mimivirus reunion Queen, Pandoravirus japonicus e Yasminevirus saudimassiliense, e foram representados como outros para preservar uma melhor visualização do gráfico devido à baixa representatividade no conjunto de dados (Figura 7B). Com relação aos demais aminoácidos, foi observado uma distribuição de 31 argininas (6,7%), 30 serinas (6,5%), 28 alaninas e cisteínas (6,0%), 27 histidinas (5,8%), 22 glutaminas, prolina, triptofanos e isoleucinas (4,7%), 21 valinas (4,5%), 20 glicinas (4,3%), 17 glutamatos (3,7%), 14 lisinas e treoninas (3,0%), 13 metioninas, tirosinas e aspartatos (2,8%), 9 asparaginas (1,9%) e 7 fenilalaninas (1,5%).

Sendo assim, arginina, serina, glutamina, prolina, glutamato, lisina, treonina, metionina, tirosina, aspartato, asparagina e fenilalanina foi mais prevalente nos

Tupanvírus. Enquanto alanina, valina e glicina foram mais prevalentes nos Pandoravírus. Restando leucina, cisteína, histidina, triptofano e isoleucina que foram mais prevalentes nos Mimivírus. Os tRNA não canônicos selenocisteína, pirrolisina e de parada foram somente em Mimivírus, Pandoravírus e Yasminevírus, respectivamente. As demais distribuições e percentuais dos grupos virais e os respectivos aminoácidos codificados podem ser visualizados na Figura 8.

	Yasminevírus	Cedratvírus	Clandestinovírus	Faustovírus	Marseillevírus	Medusavírus	Mimivírus	Mollivírus	Orpheovírus	Pacmanvírus	Pandoravírus	Tupanvírus
Leucina	5 (5,7%)	0 (0%)	0 (0%)	0 (0%)	1 (1,1%)	0 (0%)	60 (68,2%)	4 (4,5%)	0 (0%)	0 (0%)	4 (4,5%)	14 (15,9%)
Histidina	2 (7,4%)	1 (3,7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	20 (74,1%)	0 (0%)	0 (0%)	0 (0%)	2 (7,4%)	2 (7,4%)
Cisteína	2 (7,1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	20 (71,4%)	1 (3,6%)	0 (0%)	0 (0%)	2 (7,1%)	3 (10,7%)
Triptofano	2 (9,1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	12 (54,5%)	1 (4,5%)	0 (0%)	0 (0%)	4 (18,2%)	3 (13,6%)
Isoleucina	3 (13,6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	10 (45,5%)	0 (0%)	0 (0%)	2 (9,1%)	1 (4,5%)	6 (27,3%)
Metionina	2 (15,4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (15,4%)	0 (0%)	0 (0%)	3 (23,1%)	6 (46,2%)
Valina	4 (19%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	9 (42,9%)	8 (38,1%)
Serina	4 (13,3%)	0 (0%)	1 (3,3%)	3 (10%)	0 (0%)	0 (0%)	0 (0%)	1 (3,3%)	0 (0%)	0 (0%)	8 (26,7%)	13 (43,3%)
Prolina	3 (13,6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	12 (54,5%)	7 (31,8%)
Treonina	3 (21,4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (35,7%)	6 (42,9%)
Alanina	3 (10,7%)	0 (0%)	0 (0%)	0 (0%)	1 (3,6%)	0 (0%)	0 (0%)	0 (0%)	1 (3,6%)	0 (0%)	16 (57,1%)	7 (25%)
Tirosina	4 (30,8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7,7%)	0 (0%)	0 (0%)	1 (7,7%)	0 (0%)	1 (7,7%)	6 (46,2%)
Arginina	8 (25,8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (6,5%)	2 (6,5%)	0 (0%)	0 (0%)	0 (0%)	7 (22,6%)	12 (38,7%)
Glicina	6 (30%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (10%)	0 (0%)	0 (0%)	7 (35%)	5 (25%)
Glutamina	3 (13,6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (4,5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (36,4%)	10 (45,5%)
Asparagina	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (22,2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (11,1%)	6 (66,7%)
Lisina	4 (28,6%)	1 (7,1%)	0 (0%)	0 (0%)	1 (7,1%)	1 (7,1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (7,1%)	6 (47,1%)
Aspartato	4 (30,8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (23,1%)	6 (46,2%)
Glutamato	5 (29,4%)	0 (0%)	0 (0%)	0 (0%)	2 (11,8%)	0 (0%)	1 (5,9%)	0 (0%)	0 (0%)	0 (0%)	1 (5,9%)	8 (47,1%)
Fenilalanina	2 (28,6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (14,3%)	4 (57,1%)
Códon de parada	1 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Selenocisteína	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Pirrolisina	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)	0 (0%)

Figura 8. Distribuição dos aminoácidos codificados nos grupos virais.

Seguidamente, a avaliação do tamanho dos tRNA variou individualmente de 69 nt (tRNA-stop do Yasminevírus saudimassiliense) a 101 nt (tRNA-leucina do Pandoravirus japonicus) e a média dos grupos variou de 72,5 nt (Yasminevírus) a 94,5 nt (Orpheovírus). O conteúdo GC desses tRNA variou individualmente de 13%, pertencente ao tRNA-tirosina do Orpheovirus IHUMI-LCC2, a 84%, pertencente ao tRNA-prolina do Pandoravirus dulcis, e a média dos grupos variou de 22,8% (Orpheovírus) a 68,1% (Pandoravírus).

Logo, dentre os grupos de vírus analisados o Orpheovírus apresentou a maior média quanto ao tamanho dos tRNA com um valor igual a 94,5 nt (Figura 7C), ao passo que a média de seu conteúdo GC foi a menor dentre todos os grupos, com um valor igual a 22,8% (Figura 7D). Faustovírus e Marseillevírus exibiram médias de tamanho próximo ao Orpheovírus, 94 e 92,3 nt, e a média de conteúdo GC também foi próxima entre os dois grupos, 45,7% e 45,8%, respectivamente. Outros grupos que apresentaram tamanhos próximos foram os Pandoravírus (89,6 nt), também detentor da maior média de conteúdo GC dentre os grupos (68,1%), Cedratvírus (89 nt), porém

com o segundo menor conteúdo GC (42,7%), Medusavírus (87,2 nt) e Mollivírus (86,3 nt), cuja %GC se assemelhou entre os dois últimos, 63% e 62,7%, respectivamente.

Ademais, os Mimivírus, Tupanvírus, Yasminevírus e Pacmanvírus apresentaram média de tamanho de seus tRNA na mesma faixa, entre 70 nt a 80 nt, sendo 79,1 nt, 74,9 nt, 73,1 nt e 72,5 nt, respectivamente. Seguindo esse padrão, juntamente com os Marseillevírus e Faustovírus, o conteúdo GC de ambos também se apresentaram na mesma faixa, entre 40% e 50%, sendo 45,7% para Mimivírus, 50% para Tupanvírus, 48,7% para Yasminevírus e 46,9% para Pacmanvírus. Em comparação com o conteúdo GC dos genomas, o conteúdo dos tRNA, em sua maioria, não apresentou grandes diferenças. Os grupos que apresentaram uma diferença maior que 10% em relação ao conteúdo GC dos genomas para os tRNA, foram, em ordem decrescente, os Tupanvírus (20,7%), seguido pelos Mimivírus (19,1%), Clandestinovírus (15,5%) e Pacmanvírus (13,5%).

Seguidamente, a avaliação da presença de íntrons nos tRNA ocorreu na maioria dos 48 genomas utilizados nesse estudo, sendo presente em 34 e ausente em 14 vírus (Figura 9A). Desses 34, 12 são dos Mimivírus (35,3%), 8 Pandoravírus (23,5%), 3 Marseillevírus e Faustovírus (8,8%), 2 Mollivírus, Medusavírus e Cedratvírus (5,9%), e 1 Tupanvírus e Orpheovírus (3%). Os 14 vírus que não tiveram presença de íntrons ficou distribuído como 9 Mimivírus (64,3%), 2 Pacmanvírus (14,3%), e 1 Tupanvírus, Yasminevírus e Clandestinovírus (7,1%). O grupo que apresentou uma maior incidência de tRNA contendo inserções de íntrons foi o Pandoravírus, com um total de 79 íntrons (69,3%), sendo altamente discrepante com relação ao segundo grupo, o dos Mimivírus, com 12 (10,5%). Os Mollivírus, Marseillevírus e Medusavírus apresentaram o mesmo resultado para 5 tRNA (4,4%), seguido pelos Faustovírus com 3 (2,6%), Tupanvírus e Cedratvírus com 2 (1,8%) e o Orpheovírus com apenas 1 tRNA contendo íntron (0,9%) (Figura 9B).

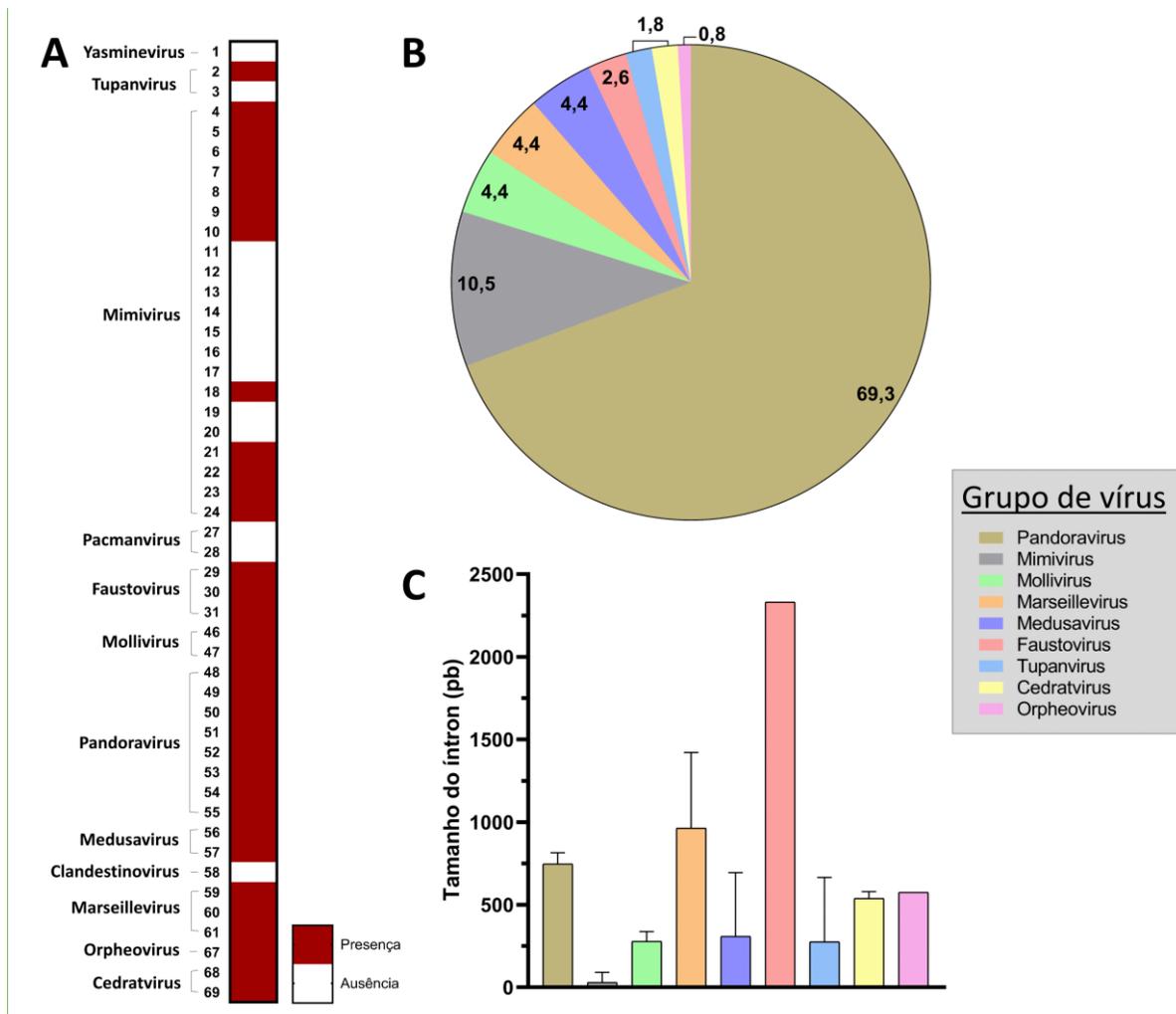


Figura 9. Presença de íntrons em tRNA virais. (A) Íntron individualmente nos vírus dos grupos de estudo; (B) Distribuição de incidência de íntrons nos grupos virais; (C) Tamanho médio dos íntrons nos grupos virais.

Quanto ao tamanho dessas regiões contendo íntrons, os Faustovírus apresentaram a maior média dentre todos, com um mesmo tamanho de seus 3 tRNA (2332 nt), o que gerou ausência de desvio padrão. (Figura 9C). Em contrapartida, o grupo dos Marseillevírus apresentou uma alta discrepância com relação ao primeiro grupo, tanto no valor médio de tamanho de íntrons, 964,3 nt, quanto no desvio padrão (variação de 120 nt a 1480 nt), que juntamente com o grupo dos Medusavírus (37 nt a 1383 nt) e Tupanvírus (6 nt a 1096 nt) foi elevado.

Os tRNA que contém inserções de íntrons apresentaram um padrão, que quando um íntron tem um valor de pares de bases igual ao de outro íntron de outro vírus do mesmo grupo, eles se apresentam iguais em número de pares de base do

tRNA, posição de inserção no tRNA e o aminoácido codificado, como exemplo os Faustovírus em que seus tRNA estão inseridos na mesma posição, entre 28/29 do tRNA e codificam para um mesmo aminoácido, o de serina. Para esse padrão houve apenas uma exceção para um íntron de tamanho 51 nt dos Pandoravírus (*Pandoravirus neocaledonia* e *Pandoravirus dulcis*) que houve divergência tanto no número de pares de base do tRNA, quando na posição de inserção do íntron e o aminoácido codificado. Esse padrão foi observado para 10 dos 11 tamanhos de íntrons iguais, sendo um mesmo valor (18 nt) para 8 dos 11 Mimivírus que apresentaram presença de íntrons, um valor (148 nt) para 2 dos 3 Marseillevírus, cinco valores (94, 116, 230, 236 e 982 nt) para 6 dos 8 Pandoravírus, sendo um íntron que não seguiu o padrão (51 nt), um valor (2332 nt) para todos 3 Faustovírus, e dois valores (299 e 341 nt) para os 2 Mollivírus. Houve uma repetição de tamanho de íntron de 52 nt entre grupo de vírus diferente (*Mimivirus* - *Megavirus chilense* e *Medusavirus* - *Medusavirus medusae*), porém quando feita essa mesma comparação de íntrons de mesmo tamanho, o padrão não se manteve, alterando tanto o número de pares de base do tRNA, posição de inserção e o aminoácido correspondente.

### 5.3 Distribuição de tRNA virais

Os grupos que apresentaram maior diversidade e quantidade de tRNA foram o Yasminevírus, com genes para 19 diferentes aminoácidos e 1 tRNA de parada, e os Tupanvírus, que apresentam genes de tRNA para os 20 aminoácidos canônicos, variando apenas na incidência destes entre o grupo. Os Mimivírus apresentaram individualmente variações quanto a quantidade de tRNA-leucina presente, enquanto para os demais aminoácidos os valores permaneceram os mesmos (1 tRNA). Quanto aos aminoácidos presentes, houve pouca variação, com uma incidência de leucina, histidina e cisteína para 20 vírus do grupo (95%), triptofano 12 (57%) e isoleucina 10 (48%). Alguns aminoácidos pontuais foram encontrados em poucos vírus pertencentes ao grupo, como arginina (*Moumouvirus moumou* e *Megavirus baoshanense*), selenocisteína (*Mimivirus reunion Queen*) e glutamato (*Moumouvirus maliensis*). Os demais aminoácidos não foram encontrados.

Por outro lado, os Pandoravírus apresentaram maior variação entre o seu grupo e os outros grupos do estudo, tanto em quantidade individual de aminoácidos como em incidência, sendo a alanina presente em 8 vírus do grupo (100%), prolina 7

(87,5%), valina, glutamina e glicina 6 (75%), serina e arginina 5 (62,5%), leucina, triptofano e treonina 3 (37,5%), histidina, metionina, cisteína e aspartato 2 (25%), e isoleucina, tirosina, asparagina, lisina, glutamato, fenilalanina e pirrolisina 1 (12,5%). Somente tRNA para selenocisteína não foi encontrado no grupo.

Além destes, os demais grupos virais, que possuem menor quantidade de integrantes, de três a dois genomas disponíveis, essa análise de incidência provoca uma discrepância alta com relação aos valores, como é possível visualizar para os Cedratvírus, que com 2 isolados, cada um com 1 tRNA para aminoácidos divergentes, gerou uma incidência de 50% para lisina, e 50% para histidina. Ao passo que, Pacmanvírus, com dois isolados, e Faustovírus, com três isolados, cada um também com apenas 1 tRNA, apresentaram somente uma mesma espécie de aminoácido no grupo, portanto ambos tiveram uma incidência de 100%, o primeiro para isoleucina, e o segundo para serina. Essa comparação não foi possível de ser feita com o Clandestinovírus, o Yasminevírus e o Orpheovírus que apresentam somente um isolado no grupo. Portanto é possível verificar apenas a distribuição e incidência dos aminoácidos codificados (Figura 10).

A maior prevalência nos Mimivírus foi de tRNA-leucina, com um resultado elevado, comparado aos outros grupos, de 60 tRNA (47,6%), o mesmo prevalente nos Mollivírus com 4 tRNA-leucina (36,4%) e os Tupanvírus com 14 tRNA-leucina (10,1%) no grupo. No Yasminevírus a prevalência foi de tRNA-arginina com 8 (11,4%), em Pandoravírus tRNA-alanina 16 (16,5%), em Marseillevírus tRNA-glutamato 2 (40%), e Pacmanvírus 2 tRNA-isoleucina (100%). Clandestinovírus e Faustovírus a prevalência foi para o mesmo tipo, tRNA-serina, sendo 1 e 3 respectivamente (100%). Os Cedratvírus e o Orpheovírus não houve prevalência de nenhum tRNA, visto que os dois grupos apresentavam dois tipos de tRNA diferentes (50%). Para os Medusavírus não houve prevalência de um tRNA, e duas espécies de tRNA apresentaram o mesmo resultado no grupo, sendo 2 argininas e 2 asparaginas (28,6%).

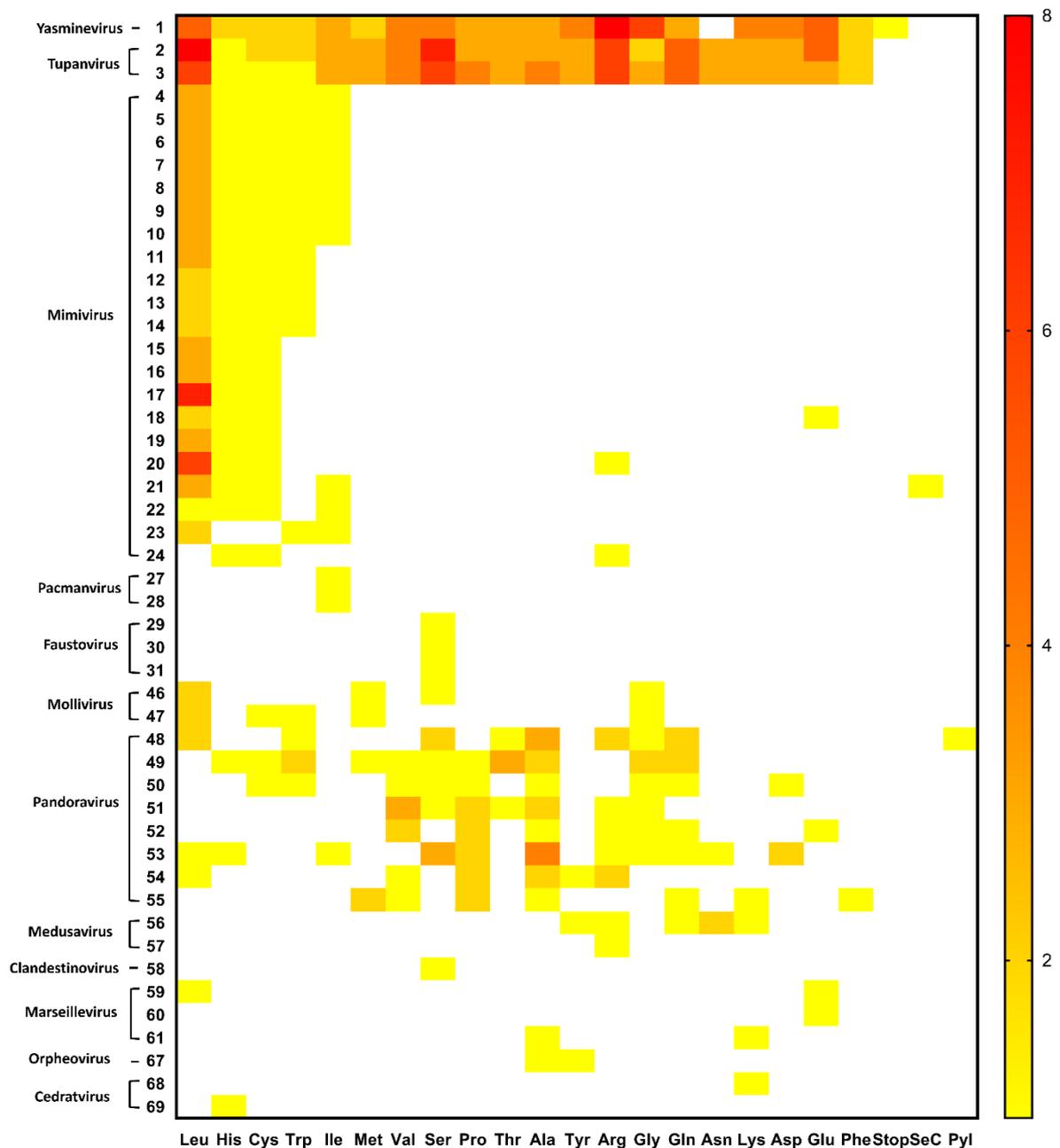


Figura 10. Heatmap da distribuição dos aminoácidos individualmente no vírus dos grupos de estudo. Espaços em branco significam ausência de tRNA no genoma viral correspondente.

#### 5.4 Localização dos tRNA nos genomas virais

Para avaliar a posição do gene de tRNA no genoma e a fita em que é expresso foram selecionados um representante de cada grupo viral que atendesse a condição

de tRNA > 1. Portanto, o grupo dos Cedratvírus e Faustovírus não foi representado nesta análise.

Logo, a maior parte dos tRNA apresentaram-se dispersos ao longo dos genomas virais. Alguns vírus como o Orpheovirus IHUMI-LCC2 (Figura 11A), Marseillevirus marseillevirus (Figura 11B) e Mollivirus sibericum (Figura 11C) exibiram os tRNA completamente dispersos. Outros como o Medusavirus medusae (Figura 11D), Mimivirus bradfordmassiliense (Figura 12A) e Moumouvirus moumou (Figura 12B) apresentaram apenas uma região com agrupamento de dois tRNA, com uma diferença de 1 a 47 nt entre o término de um tRNA e início do outro. Os demais também estavam dispersos ao longo do genoma. Os dois isolados do grupo dos Mimivírus apresentaram um padrão de localização nessa região, sendo composta pelos tRNA-histidina e tRNA-cisteína. Porém, no caso do Moumouvirus moumou, essa região se revelou presente próximo a  $\pm 96$  Mbp no genoma, enquanto para o Mimivirus bradfordmassiliense esteve próximo de  $\pm 350$  Mbp no genoma. O Megavirus chilense (Figura 12C) apresentou apenas uma região com 2 tRNA-leucinas relativamente próximos (1054 nt), enquanto os demais estavam completamente dispersos.

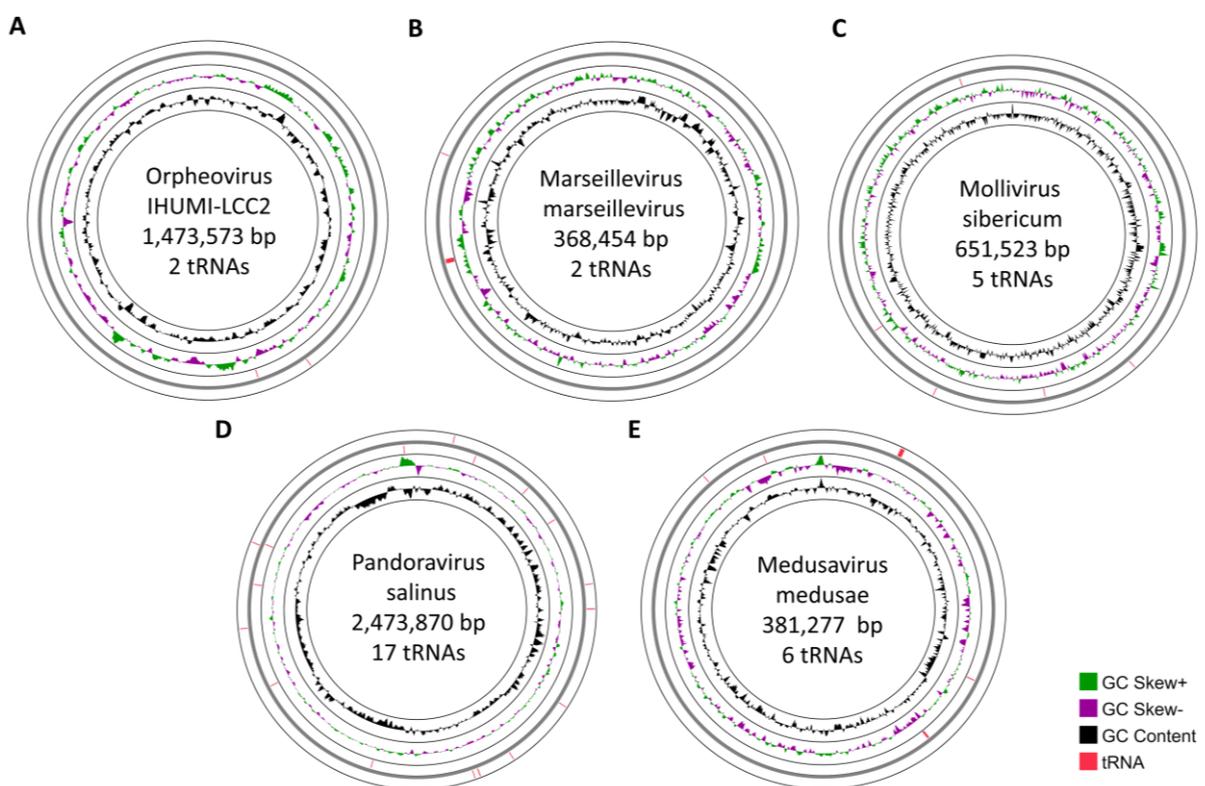


Figura 11. Mapa circular gráfico do genoma dos representantes dos grupos virais ao centro o nome do vírus, tamanho do genoma e quantidade de tRNA

presentes. De fora para o centro, o anel 1 mostra os genes de tRNA na fita direta e o anel 2 na fita reversa; o anel 3 mostra o conteúdo G+C% e o anel 4 mostra o desvio GC; (A) Orpheovírus; (B) Marseillevírus; (C) Mollivírus; (D) Pandoravírus; (E) Medusavírus.

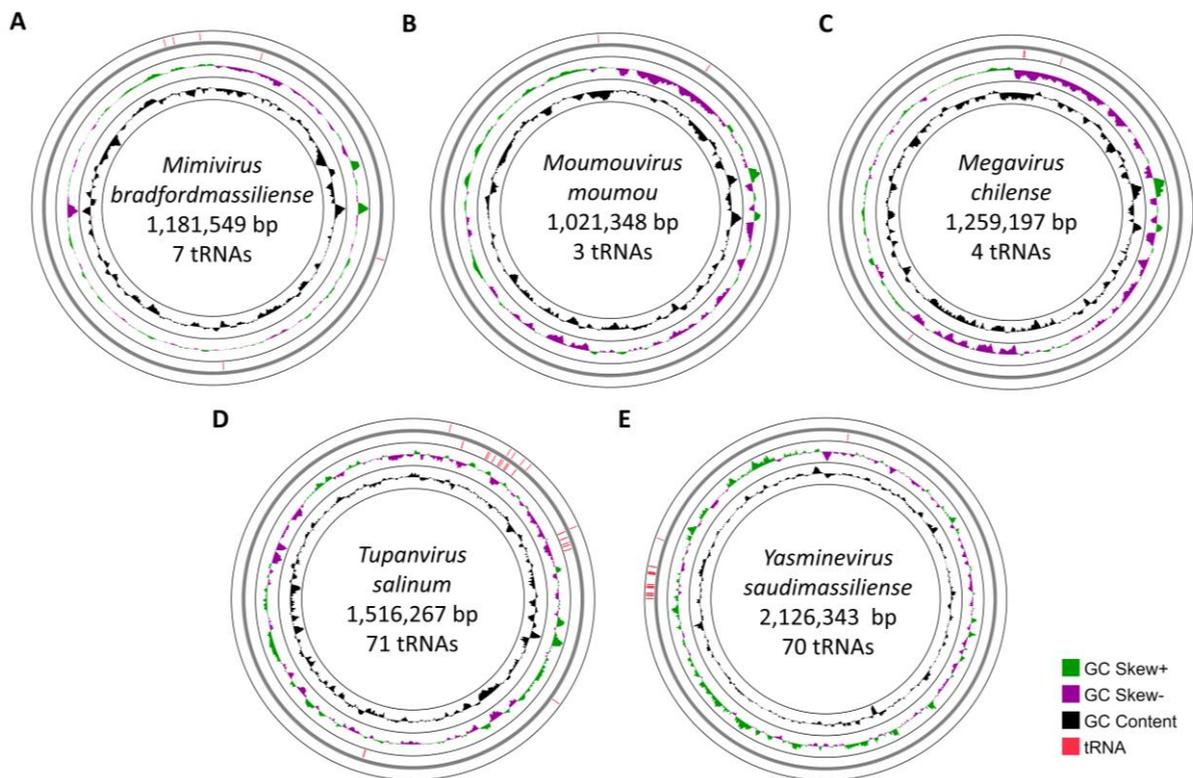


Figura 12. Mapa circular gráfico do genoma dos representantes dos grupos virais da ordem dos *Immitervirales*. Ao centro o nome do vírus, tamanho do genoma e quantidade de tRNA presentes. De fora para o centro, o anel 1 mostra os genes de tRNA na fita direta e o anel 2 na fita reversa; o anel 3 mostra o conteúdo G+C% e o anel 4 mostra o desvio GC; (A) Mimivírus; (B) Mousmouvírus; (C) Megavírus; (D) Tupanviro; (E) Yasmineviro.

Em contrapartida, esse cenário de dispersão dos tRNA ao longo do genoma muda para os vírus que codificaram um elevado número desses genes, como é para o *Tupanvirus salinum* (Figura 12D), onde aproximadamente 96% estão agrupados, e 97,2% dos tRNA totais localizados no primeiro quarto do genoma, e o *Yasminevirus saudimassiliense* (Figura 12E), sendo aproximadamente 97% de seus tRNA encontrados em tandem, e 98,6% localizados no último quarto do genoma. O *Tupanvirus salinum* apresentou apenas 3 tRNA isolados no genoma, os demais

estavam em grupos de 2 a 8 tRNA, com espaçamento de 2066 a 301.070 bp entre o término de um agrupamento e início de outro (Figura 11E).

Ainda, alguns tipos de tRNA presentes em um número >1 no genoma mostraram uma tendência a serem codificados na mesma região que seus semelhantes, como é o caso da tirosina, fenilalanina, metionina, e a maioria das leucinas e dos glutamatos. Ainda foi observado que alguns tipos de tRNA pareciam se associar a outros, como a glutamina e a serina que estavam juntas em duas regiões diferentes no genoma, observado também para duas das isoleucinas e lisinas.

Enquanto o Yasminevirus saudimassiliense apresentou apenas 2 tRNA isolados no genoma, os demais estavam em grupos de 2 a 10, com o espaçamento entre o término de um agrupamento e início de outro de 904 a 18.047 bp. Esses grupos foram encontrados próximos, entre aproximadamente 1494 kbp e 1564 kbp do genoma. Também apresentaram alguns tipos de tRNA codificados na mesma região que seus semelhantes, como foi para a prolina, cisteína, isoleucina, e a maioria das alaninas e valinas. Assim como em Tupanvírus, a glutamina e a serina também estavam expressas em uma mesma região no Yasminevírus.

### **5.5 Caracterização molecular de tRNA-leucina em vírus gigantes**

Os tRNA-leucina variaram de 73 nt, pertencente ao Yasminevirus saudimassiliense, e 101 nt, pertencente ao Pandoravirus japonicus. Os 88 tRNA-leucina presentes nos vírus gigantes apresentaram um tamanho médio de aproximadamente 84 nt e conteúdo GC de 51%, e em apenas 8 tRNA havia inserções de íntrons de tamanhos que variaram de 6 nt (Tupanvirus salinum – anticódon TAA, posição no genoma 286334, 286410) a 1047 nt (Pandoravirus macleodensis – anticódon GAG, posição no genoma 1751997, 1752854), gerando uma média de 395,6 nt e alto desvio padrão de  $\pm 319,5$  nt. Ainda foi possível observar que os íntrons dos tRNA-leucina de anticódon GAG, presente nos Mollivirus siberium, Mollivirus kamchatka e Pandoravirus dulcis, estavam inseridos na posição 35/36 do tRNA. O anticódon CAG ocorreu por duas vezes em Pandoravirus japonicus, porém os íntrons estavam inseridos em posições diferentes no tRNA. Quanto aos outros anticódons, não houve uma recorrência maior que um para se fazer um comparativo da posição de inserção.

Foram escolhidos um representante de cada grupo viral para comparar as seqüências de nucleotídeos para o gene de tRNA-leucina, o qual foi o mais frequente, comum a 27 dos 46 vírus do estudo. Como o grupo dos Pacmanvírus, Faustovírus, Medusavírus, Clandestinovírus, Orpheovírus e Cedratvírus não apresentaram nenhum gene para tRNA-leucina, esses foram desconsiderados na análise. Foram incluídos todos os anticódons de tRNA codificados por cada representante, gerando uma variação de 1 a 8 seqüências para os vírus selecionados e uma diversidade de 6 espécies diferentes.

As seqüências para o gene codificador da tRNA-leucina foram muito diversos entre os vírus analisados, sendo que em apenas 2 sítios (99 e 107) o nucleotídeo foi igual para todas as 30 seqüências. Os demais variaram, indicando que há pouca conservação nestes genes (Figura 13). Quando analisado por anticódon entre os representantes, também há pouca conservação. O mesmo acontece com as seqüências de tRNA-leucina dentro dos próprios grupos virais, incluindo todos os anticódons, com exceção do grupo dos Mimivírus. Porém, quando analisado essas mesmas seqüências entre o próprio grupo, levando em consideração os tipos de anticódons, foi possível observar, em sua maioria, uma alta taxa de conservação.

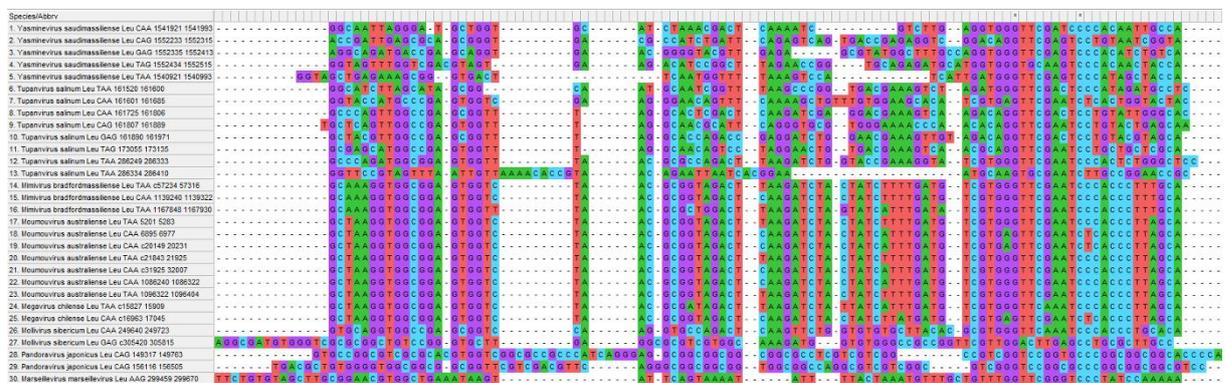


Figura 13. Alinhamento das seqüências de leucina dos representantes dos grupos virais selecionados com os respectivos anticódons da seqüência correspondente e a posição em que é expresso no genoma.

Com um total de 60 seqüências para os tRNA-leucina, os Mimivírus apresentaram todos os seus tRNA de tamanho 83 nt, do qual em 14 sítios houve variação dos nucleotídeos (83,1% de conservação) para os 3 clados, variando entre dois nucleotídeos ( $A \leftrightarrow T$ ;  $T \leftrightarrow C$ ;  $G \leftrightarrow C$ ;  $A \leftrightarrow G$ ), e em apenas dois sítios, na posição 28 e 44 da seqüência, houve a variação de três nucleotídeos ( $C \leftrightarrow G \leftrightarrow T$ ;  $C \leftrightarrow G \leftrightarrow A$ )





(94% de conservação), entre o 16 e o 51, e apresentaram 3 tipos de sequências (sequências 1 – 2, 3 – 6, 7 – 8) (Figura 14C).

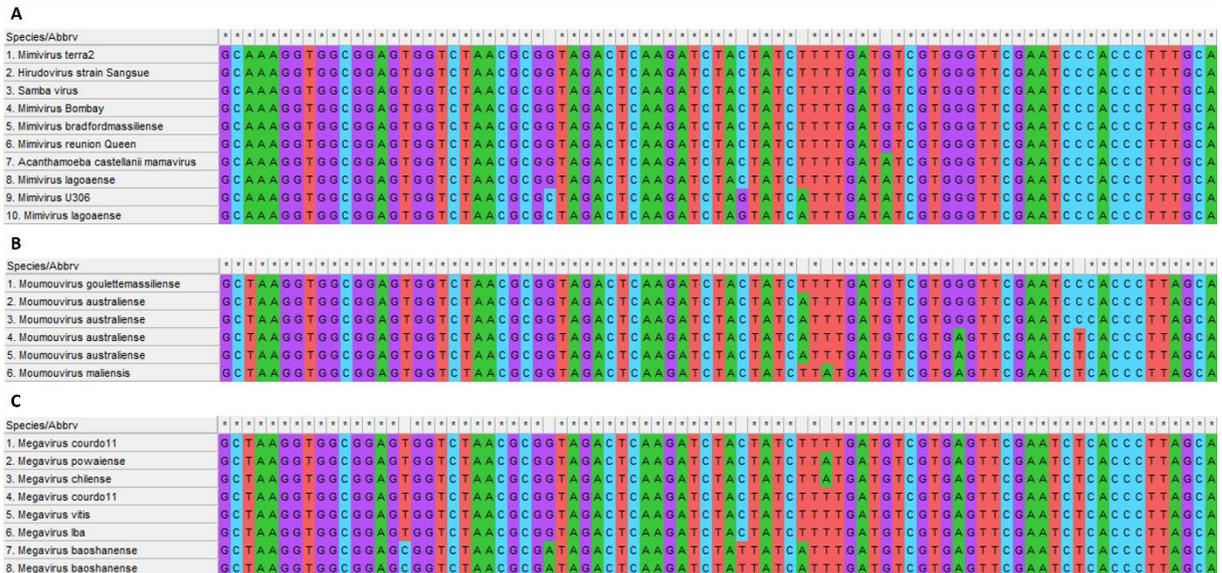


Figura 14. Alinhamento das sequências de leucina dos Mimivírus anticódon CAA. (A) Clado A – Mimivirus; (B) Clado B – Moutmovirus; (C) Clado C – Megavirus.

Os Tupanvírus apresentaram um total de 14 sequências, sendo 5 tRNA-leucina<sup>TAA</sup> de tamanho entre 77 e







## 6. DISCUSSÃO

O genoma dos vírus gigantes de DNA impressiona pela sua extensão e presença de genes relacionados ao processo de tradução. Bailly-Bechet et al. (2007) analisaram um conjunto de fagos e concluíram que a presença de genes para tRNA estava relacionada ao comprimento do genoma. Morgado et al. (2019) analisaram um conjunto de vírus em larga escala e concluíram o mesmo viés, enfatizando a correlação entre o número de genes tRNA e o comprimento do genoma. Porém, para os vírus gigantes essa tendência não foi observada. Os Tupanvírus que apresentam o maior número de tRNA, até 72, possuem um genoma de até 1,51 Mbp, enquanto, os maiores genomas com até 2,5 Mbp pertencentes ao Pandoravírus, codificam até 18 tRNA (ABRAHÃO et al., 2018; PHILIPPE et al., 2013).

Além disso, Morgado et al. (2019) também discute que os vírus com genomas menores tendem a uma organização dos tRNA em clusters para favorecer a compactação do genoma. Neste sentido, vírus gigantes não teriam a mesma tendência e esses genes estariam dispersos pelo genoma. Diferentemente da correlação positiva de comprimento do genoma e número de tRNA expressos, o agrupamento desses genes ocorre de forma inversa ao tamanho do genoma. Ainda, foi relatado nesse mesmo estudo uma exceção nos Tupanvírus que carregam de 10 a 11 genes em pequenos grupos, enquanto os demais vírus gigantes apresentaram os seus genes de tRNA dispersos homoganeamente no genoma. Essas observações são condizentes com os resultados do estudo anterior, porém, foi observado que os Tupanvírus carregam até 8 genes em agrupamentos, sendo a maior incidência entre 4 e 6 genes. E, assim como em Tupanvírus, em Yasminevírus, isolado em 2020, também foram observados tRNA organizados em grupos.

Ademais, Duncan et al. (2020) indica que com exceção dos vírus da família *Mimiviridae*, a maioria dos outros grandes vírus de DNA possui poucos ou nenhum gene para tRNA, e quando um vírus codifica a partir de dois tRNA eles não são agrupados, enfatizando os achados deste estudo. Porém, é importante salientar que com exceção dos Tupanvírus e Yasminevírus que apresentaram diversos genes agrupados, outros vírus da família *Mimiviridae* apresentaram apenas um agrupamento com dois tRNA, os demais estavam dispersos ao longo do genoma. Sendo assim, a proporção de genes dispersos é maior que os organizados em agrupamentos dentro do grupo dos vírus gigantes de amebas.

Foi evidenciado que os tRNA exercem outras funções nas células de eucariotos e procariotos (LYONS; FAY; IVANOV, 2018). Além disso, conforme descrito para fagos, genes de tRNA encontrados em genomas de vírus sugerem que eles desempenham não somente um papel traducional, mas também um papel não canônico na sua biologia na regulação da tradução, empacotamento e iniciação da transcrição reversa (ALBERS; CZECH, 2016). Portanto, atribuir aos genes de tRNA virais apenas a função na síntese de proteínas parece limitado, e por não estar claro o papel desses tRNA nos genomas de vírus gigantes, faz-se necessário um estudo mais aprofundado para se verificar a hipótese de papéis não canônicos exercidos pelos genes de tRNA desses genomas.

Quanto a predição dos tRNA pelo ARAGORN, o Clandestinovírus apresentou apenas um gene, sendo um tRNA-serina. No seu estudo original foi descrito, além do tRNA-serina, também um pseudogene de tRNA com similaridade de 32% para tRNA-histidina (ROLLAND et al., 2021). Curiosamente, foi relatado que os genes de tRNA presentes nos genomas de vírus gigantes não são pseudogenes, pois são expressos durante o ciclo de replicação deste vírus, sendo classificados em três classes de transcritos conforme o momento de expressão: precoce, intermediário e tardio (SILVA et al., 2015). Mas, quanto ao pseudogene encontrado em Clandestinovírus, nenhuma análise foi feita para verificar sua participação no ciclo de replicação. Portanto, a presença do pseudo tRNA pode indicar uma função canônica comprometida, estar relacionado a outras funções, ou ainda representar uma classificação incorreta de alguns genes de tRNA putativos de baixa pontuação pelo programa, provavelmente canônicos, como pseudogenes (LEGENDRE et al., 2010; VAN DER GULIK et al., 2023).

Ainda na predição dos tRNA, o Tupanvirus salinum, Yasminevirus saudimassiliense e Medusavirus medusae apresentaram genes de tRNA cujo isotipo e o anticódon previsto por ARAGORN não foram discriminados com precisão, sendo denominados como tRNA ambíguos. Nesses casos, o software indicou dois isotipos para um único gene tRNA, sendo um resultado inconsistente. Uma reanálise feita por este estudo utilizando o programa tRNA-scanSE 2.0 também retornou um tipo indeterminado. Um retorno pelo programa de um tRNA ambíguo é resultado da degeneração do códon que consiste na capacidade de um aminoácido ser codificado por mais de um códon (GONZALEZ; GIANNERINI; ROSA, 2019; OU et al., 2019). Portanto, na ausência de definição de ao menos um nucleotídeo que compõe o

anticódon, não é possível definir com exatidão qual o aminoácido codificado, gerando uma ambiguidade.

É estabelecido que os RNA de transferência (tRNA) variam de comprimento entre 70 e 100 nucleotídeos (KRAHN; FISCHER; SÖLL, 2020). E realmente os tRNA identificados nos genomas de vírus gigantes incluídos neste estudo revelaram esse intervalo de comprimento, porém, houve em 3 tRNA um comprimento fora desse padrão, sendo dois genes tRNA-leucina de comprimento de 101 nt em Pandoravírus e um gene tRNA de parada (anticódon TTA) de comprimento de 69 nt em Yasminevírus. A presença de um tRNA de parada em Yasminevírus foi predita pelo ARAGORN e confirmada por tRNA-scanSE 2.0, na sequência depositada no GenBank e no artigo original, mas nenhuma discussão foi feita sobre esse achado. A ocorrência desse gene é atípica, pois não foi descrito anteriormente para outros vírus, ou mesmo para organismos celulares. Curiosamente, foi relatada a presença desse mesmo tRNA em um virófago e foi levantada a hipótese de ser um indicativo de reatribuição de códon de parada nos hospedeiros desse virófago (PAEZ-ESPINO et al., 2019). Portanto são necessários maiores estudos para confirmar do que se tratar efetivamente desse tRNA de parada e descartar a possibilidade de um resultado falso positivo decorrente do algoritmo, bem como, em caso positivo, análises para confirmar o papel destes genes durante o ciclo de replicação destes vírus.

O número e a posição dos íntrons no tRNA variam dependendo do organismo (LOPES et al., 2015). Os vírus gigantes, em sua grande maioria, foram capazes de superar e muito o intervalo conhecido de 6 a 133 nt para íntrons eucarióticos, apresentando uma média de 584,5. Vários genes de tRNA baculoviral também exibiram longos íntrons, diferentemente dos chlorovirus que seguem o padrão de intervalo conhecido (DUNCAN; DUNIGAN; VAN ETTEN, 2020; OLIVEIRA et al., 2022).

Em *Mimivírus*, todos os seus íntrons se apresentavam inseridos na posição canônica 37/38. A inserção na posição canônica interrompe a estrutura do anticódon, dividindo o tRNA em metades não funcionais, fazendo com que seu splicing seja indispensável. Os demais vírus apresentaram os íntrons inseridos mais em posições não canônicas, entre 27/28 e 41/42, posições que não parecem interromper a estrutura geral do tRNA. Os íntrons encontrados em genes de tRNA eucarióticos e arqueais são inseridos principalmente na posição 37/38, também conhecida como posição canônica de íntrons emendados enzimaticamente nos precursores de tRNA,

mas também podem ser encontrados inseridos em posições não canônicas, sendo encontrado em arqueias uma maior diversidade de posições de inserções de íntrons em comparação aos eucariotos. Além disso, podem apresentar em ambos mais de uma sequência intrônica no mesmo gene de tRNA, característica que não foi observada em nenhum vírus com presença de sequência intrônica. A origem dos íntrons de tRNA na posição canônica é considerada ancestral e a variabilidade posicional reflete nas mudanças de especificidade enzimática para splicing, que presumivelmente é o resultado da coevolução entre os íntrons e essas enzimas (FUJISHIMA; KANAI, 2014; LOPES et al., 2015; SUGAHARA et al., 2009; YOSHIHISA, 2014)

Além disso, os genes de tRNA com inserção de íntrons que apresentaram as mesmas características (posição do íntron no gene, tamanho e aminoácido codificado) foram analisados para se verificar a conservação e se poderia indicar uma duplicação de genes. Em Faustovírus, os genes foram idênticos, enquanto em Pandoravírus, 3/5 foram 100% idênticos (tamanho 230, 236 e 982 nt). O íntron de tamanho 116 nt foi semelhante para dois dos três vírus (Pandoravirus quercus e Pandoravirus celtis), e divergente em 17 sítios, nove no íntron e oito no tRNA, de Pandoravirus Inopinatum, enquanto o de comprimento de 94 nt divergiu em 12 sítios, sendo apenas cinco no íntron. Em Mimivírus também foram idênticos, com exceção do Mimivirus terra2 que divergiu do grupo em cinco sítios, todos no tRNA e não no íntron, e o Acanthamoeba polyphaga mimivirus Oyster em um único nucleotídeo, também no tRNA. Os íntrons dos Mollivírus, comprimento 299 e 341 nt, apresentaram 4 e 5 nucleotídeos diferentes, sendo três localizados no íntron e apenas um no tRNA, e os cinco também no íntron, respectivamente. O íntron do Marseillévirus, comprimento 1574 nt apresentou 12 sítios diferentes e todos no íntron.

Diante disso é possível observar que há a presença de uma possível duplicação de genes para algumas sequências em grupos de vírus que mantiveram os fragmentos completamente conservados. Micheli et al. (2022) infere que a associação de íntrons com a duplicação de genes aumenta a estabilidade das cópias duplicadas e ambos os processos são suportados evolutivamente. Ao mesmo tempo é curioso pensar que longos íntrons apresentaram nenhuma ou poucas alterações na sequência visto que não interferem diretamente na função do gene, pois não são codificantes, e tendem a acumular mais mutações. Muito pouco se sabe sobre o destino e a função

dos íntrons de tRNA, entretanto é provável que haja uma vantagem seletiva em transportar esses íntrons dentro do genoma (CHOREV; CARMEL, 2012).

Não somente nos tRNA, a ocorrência de íntrons também foi relatada na proteína principal do capsídeo em Faustovírus (com até 18 íntrons presentes em Faustovirus S17) e Mimivirus (Mimivirus bradfordmassiliense consistindo em dois íntrons, e Acanthamoeba castellanii mamavirus consistindo em um íntron). Introns também foram detectados em outros genes de vírus gigantes de amebas, como em genes de polimerase de DNA. A presença de íntrons sugere que há uma influência de modo a aumentar a expressão gênica em eucariotos, e a presença de splicing descrita em outros vírus pode reforçar essa hipótese de aumento de expressão gênica (AZZA et al., 2009; COLSON et al., 2011; PHILIPPE et al., 2013; RETENO et al., 2015).

Os quatro genes de tRNA que foram mais frequentes individualmente entre os vírus (tRNA-leucina, tRNA-histidina, tRNA-cisteína e tRNA-triptofano) foram analisados por análise filogenética (dados não mostrados), onde os íntrons presentes nos tRNA que o continham foram removidos e as árvores construídas pelo método de máxima verossimilhança, mas devido à baixa qualidade dos resultados, justificado por sequências muito divergentes e/ou sequências muito semelhantes, foram descartados do estudo e a filogenia a partir das sequências de tRNA não foi possível de ser construída.

Widmann et al. (2010) descreve os tRNA como candidatos não favoráveis a estudos filogenéticos dado ao tamanho reduzido das sequências. Para isso, alguns estudos utilizam de genes concatenados, em que as sequências individuais dos tRNA são conectadas em um único alinhamento para produzir sequências com maior quantidade de dados. Porém, ao se utilizar genes concatenados pode-se induzir positivamente um erro na inferência de sua origem (EDWARDS; LIU; PEARL, 2007; MORGADO; VICENTE, 2019a; WANG et al., 2021). Ademais, a utilização de sequências individuais de tRNA com íntrons podem induzir a erros na montagem da árvore que por acumularem mais mutações acarretam altas taxas de divergência entre as sequências (SIMON et al., 2003).

Por outro lado, como foi observado nesse estudo, os íntrons de vírus gigantes apresentaram em sua maioria pouca ou nenhuma variação em suas sequências, o que torna a filogenia de tRNA com íntrons uma possível metodologia futura mais eficiente para se obter uma árvore com maior resolução. Portanto são necessários outros estudos para construção da filogenia dos tRNA virais.

## 7. CONCLUSÃO

Das setenta sequências depositadas no GenBank, quarenta e oito continham ao menos um gene de tRNA. Essa distribuição se concentrou principalmente nos Tupanvírus, que são os vírus com maior quantidade de genes tRNA presentes, e nos Mimivírus, grupo com um grande número de isolados. Encontrou-se grande diversidade na identidade do anticódon, mas foi observado uma tendência entre os vírus para aqueles correspondentes principalmente ao aminoácido leucina. A organização desses tRNA em agrupamentos ocorreu em vírus com elevada quantidade de tRNA codificados, como o Tupanvírus e o Yasminevírus. Os íntrons contidos nesses genes também foram avaliados e fizeram-se presentes em trinta e quatro genomas. Não foi possível traçar uma árvore filogenética dos tRNA em vírus gigantes, mas pode-se observar que, em sua maioria, as sequências apresentaram alta taxa de conservação entre os membros dos grupos ao qual pertencem.

## 8. REFERÊNCIAS

- ABERGEL, C.; LEGENDRE, M.; CLAVERIE, J. M. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. **FEMS Microbiology Reviews**, v. 39, n. 6, p. 779–796, 1 nov. 2015.
- ABRAHÃO, J. et al. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. **Nature Communications**, v. 9, n. 1, 1 dez. 2018.
- ALBERS, S.; CZECH, A. Exploiting tRNAs to Boost Virulence. **Life**, v. 6, n. 1, 19 jan. 2016.
- ANDREANI, J. et al. Cedratvirus, a Double-Cork Structured Giant Virus, is a Distant Relative of Pithoviruses. **Viruses**, v. 8, n. 11, 1 nov. 2016.
- ANDREANI, J. et al. Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads between Asfarviridae and Faustoviruses. **Journal of virology**, v. 91, n. 14, 15 jul. 2017.
- ANDREANI, J. et al. Orpheovirus IHUMI-LCC2: A new virus among the giant viruses. **Frontiers in Microbiology**, v. 8, n. JAN, 22 jan. 2018.
- ARANTES, T. S. et al. The Large Marseillevirus Explores Different Entry Pathways by Forming Giant Infectious Vesicles. **Journal of Virology**, v. 90, n. 11, p. 5246, 6 jun. 2016.
- AYLWARD, F. O. et al. A phylogenomic framework for charting the diversity and evolution of giant viruses. **PLOS Biology**, v. 19, n. 10, p. e3001430, 1 out. 2021.
- AZZA, S. et al. Revised Mimivirus major capsid protein sequence reveals intron-containing gene structure and extra domain. **BMC Molecular Biology**, v. 10, p. 39, 11 maio 2009.
- BAILLY-BECHET, M.; VERGASSOLA, M.; ROCHA, E. Causes for the intriguing presence of tRNAs in phages. **Genome Research**, v. 17, n. 10, p. 1486, out. 2007.
- BAJRAI, L. H. et al. Kaumobavirus, a New Virus That Clusters with Faustoviruses and Asfarviridae. **Viruses**, v. 8, n. 11, 1 nov. 2016.
- BAJRAI, L. H. et al. Isolation of Yasminevirus, the First Member of Klosneuvirinae Isolated in Coculture with *Vermamoeba vermiformis*, Demonstrates an Extended Arsenal of Translational Apparatus Components. **Journal of Virology**, v. 94, n. 1, 12 dez. 2020.
- BAR-ON, Y. M.; PHILLIPS, R.; MILO, R. The biomass distribution on Earth. **Proceedings of the National Academy of Sciences of the United States of America**, v. 115, n. 25, p. 6506–6511, 19 jun. 2018.
- BOS, L. Hundred years of Koch's Postulates and the history of etiology in plant virus research. **Netherlands Journal of Plant Pathology**, v. 87, n. 3, p. 91–110, 1981.

- BOS, L. Beijerinck's work on tobacco mosaic virus: historical context and legacy. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 354, n. 1383, p. 675, 3 mar. 1999.
- BOYER, M. et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 51, p. 21848, 12 dez. 2009.
- CHOREV, M.; CARMEL, L. The function of introns. **Frontiers in Genetics**, v. 3, n. APR, p. 24171, 13 abr. 2012.
- CLAVERIE, J. M. et al. Mimivirus and Mimiviridae: Giant viruses with an increasing number of potential hosts, including corals and sponges. **Journal of Invertebrate Pathology**, v. 101, n. 3, p. 172–180, 1 jul. 2009.
- COLSON, P. et al. Viruses with More Than 1,000 Genes: Mamavirus, a New Acanthamoeba polyphaga mimivirus Strain, and Reannotation of Mimivirus Genes. **Genome Biology and Evolution**, v. 3, n. 1, p. 737, 2011.
- COLSON, P. et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. **Archives of virology**, v. 158, n. 12, p. 2517, dez. 2013.
- CREAGER, A. N. H. Tobacco Mosaic Virus and the History of Molecular Biology. <https://doi.org/10.1146/annurev-virology-100520-014520>, v. 9, p. 39–55, 29 set. 2022.
- DELESALLE, V. A. et al. Testing hypotheses for the presence of tRNA genes in mycobacteriophage genomes. **Bacteriophage**, v. 6, n. 3, p. e1219441, 2 jul. 2016.
- DUNCAN, G. A.; DUNIGAN, D. D.; VAN ET TEN, J. L. Diversity of tRNA Clusters in the Chloroviruses. **Viruses**, v. 12, n. 10, 16 out. 2020.
- EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 1 mar. 2004.
- EDWARDS, S. V.; LIU, L.; PEARL, D. K. High-resolution species trees without concatenation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 14, p. 5936–5941, 3 abr. 2007.
- FRAENKEL-CONRAT, H. Tobacco Mosaic Virus The History of Tobacco Mosaic Virus and the Evolution of Molecular Biology. **The Plant Viruses**, p. 5–17, 1986.
- FUJISHIMA, K.; KANAI, A. tRNA gene diversity in the three domains of life. **Frontiers in Genetics**, v. 5, n. MAY, 2014.
- GONZALEZ, D. L.; GIANNERINI, S.; ROSA, R. On the origin of degeneracy in the genetic code. **Interface Focus**, v. 9, n. 6, 6 dez. 2019.
- HENDRIX, R. W. et al. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 5, p. 2192, 3 mar. 1999.

ICTV. **Taxonomy Release History.** Disponível em: <<https://ictv.global/taxonomy/history>>.

ICTV. **Current ICTV Taxonomy Release.** Disponível em: <<https://ictv.global/taxonomy>>.

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

JAHN, D.; VERKAMP, E.; SÖLL, D. Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. **Trends in Biochemical Sciences**, v. 17, n. 6, p. 215–218, 1 jun. 1992.

KOONIN, E. V. et al. Global Organization and Proposed Megataxonomy of the Virus World. **Microbiology and Molecular Biology Reviews : MMBR**, v. 84, n. 2, 20 maio 2020.

KOONIN, E. V.; YUTIN, N. Origin and Evolution of Eukaryotic Large Nucleocytoplasmic DNA Viruses. **Intervirology**, v. 53, n. 5, p. 284, jun. 2010.

KOONIN, E. V.; YUTIN, N. Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. Em: **Advances in Virus Research**. [s.l.] Academic Press Inc., 2019. v. 103p. 167–202.

KRAHN, N.; FISCHER, J. T.; SÖLL, D. **Naturally Occurring tRNAs With Non-canonical Structures**. **Frontiers in Microbiology** Frontiers Media S.A., , 21 out. 2020.

LA SCOLA, B. et al. A giant virus in amoebae. **Science**, v. 299, n. 5615, p. 2033, 28 mar. 2003.

LA SCOLA, B. , et al. **Eighth Report of the International Committee on Taxonomy of Viruses**. 1. ed. Amsterdam, The Netherlands: [s.n.].

LASLETT, D.; CANBACK, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. **Nucleic Acids Research**, v. 32, n. 1, p. 11, 2004.

LECOQ, H. Découverte du premier virus, le virus de la mosaïque du tabac : 1892 ou 1898 ? **Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie**, v. 324, n. 10, p. 929–933, 1 out. 2001.

LEGENDRE, M. et al. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. **Genome Research**, v. 20, n. 5, p. 664, 2010.

LEGENDRE, M. et al. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 11, p. 4274–4279, 18 mar. 2014.

LEGENDRE, M. et al. In-depth study of Mollivirus sibericum, a new 30,000-yold giant virus infecting Acanthamoeba. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 38, p. E5327–E5335, 22 set. 2015.

- LOPES, R. R. S. et al. Cutting, dicing, healing and sealing: the molecular surgery of tRNA. **Wiley interdisciplinary reviews. RNA**, v. 6, n. 3, p. 337, 1 maio 2015.
- LOUTEN, J. Virus Structure and Classification. **Essential Human Virology**, p. 19, 2016.
- LWOFF, A. The concept of virus. **Journal of general microbiology**, v. 17, n. 2, p. 239–253, 1 out. 1957.
- LYONS, S. M.; FAY, M. M.; IVANOV, P. The role of RNA modifications in the regulation of tRNA cleavage. **FEBS letters**, v. 592, n. 17, p. 2828, 1 set. 2018.
- MARQUET, R. et al. tRNAs as primer of reverse transcriptases. **Biochimie**, v. 77, n. 1–2, p. 113–124, 1995.
- MAYER, K. A. et al. Hijacking the supplies: Metabolism as a novel facet of virus-host interaction. **Frontiers in Immunology**, v. 10, n. JULY, p. 1533, 3 jul. 2019.
- MICHELI, G.; CAMILLONI, G. Can Introns Stabilize Gene Duplication? **Biology**, v. 11, n. 6, 1 jun. 2022.
- MÖNTTINEN, H. A. M. et al. The genomes of nucleocytoplasmic large DNA viruses: viral evolution writ large. **Microbial Genomics**, v. 7, n. 9, p. 649, 2021.
- MORENO-ALTAMIRANO, M. M. B.; KOLSTOE, S. E.; SÁNCHEZ-GARCÍA, F. J. Virus Control of Cell Metabolism for Replication and Evasion of Host Immune Responses. **Frontiers in Cellular and Infection Microbiology**, v. 9, n. APR, p. 95, 2019.
- MORGADO, S. M.; VICENTE, A. C. P. Exploring tRNA gene cluster in archaea. **Memórias do Instituto Oswaldo Cruz**, v. 114, n. 1, p. 1–6, 1 jan. 2019a.
- MORGADO, S.; VICENTE, A. C. Global In-Silico Scenario of tRNA Genes and Their Organization in Virus Genomes. **Viruses**, v. 11, n. 2, 1 fev. 2019b.
- OLIVEIRA, H. DE P. et al. Identification and analysis of putative tRNA genes in baculovirus genomes. **Virus Research**, v. 322, p. 198949, 1 dez. 2022.
- OU, X. et al. Errors in translational decoding: tRNA wobbling or misincorporation? **PLOS Genetics**, v. 15, n. 3, p. e1008017, 1 mar. 2019.
- PAEZ-ESPINO, D. et al. Diversity, evolution, and classification of virophages uncovered through global metagenomics. **Microbiome**, v. 7, n. 1, p. 1–14, 10 dez. 2019.
- PARVATHY, S. T.; UDAYASURIYAN, V.; BHADANA, V. Codon usage bias. **Molecular Biology Reports**, v. 49, n. 1, p. 539, 1 jan. 2022.
- PHILIPPE, N. et al. Pandoraviruses: Amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. **Science**, v. 341, n. 6143, p. 281–286, 2013.
- RAINA, M.; IBBA, M. tRNAs as regulators of biological processes. **Frontiers in Genetics**, v. 5, n. JUN, 2014.

- RAOULT, D. et al. The 1.2-megabase genome sequence of Mimivirus. **Science (New York, N.Y.)**, v. 306, n. 5700, p. 1344–1350, 19 nov. 2004.
- RAOULT, D.; FORTERRE, P. Redefining viruses: lessons from Mimivirus. **Nature Reviews Microbiology** **2008 6:4**, v. 6, n. 4, p. 315–319, 3 mar. 2008.
- RETENO, D. G. et al. Faustovirus, an Asfarvirus-Related New Lineage of Giant Viruses Infecting Amoebae. **Journal of Virology**, v. 89, n. 13, p. 6585, 7 jul. 2015.
- ROBERT W. HOLLEY et al. Structure of a Ribonucleic Acid. **Science**, v. 147, n. 3664, p. 1462–1465, 19 mar. 1965.
- RODRIGUES, R. A. L. et al. Giant Viruses and Their Virophage Parasites. Em: **Encyclopedia of Virology**. [s.l.] Elsevier, 2021. p. 372–381.
- ROLLAND, C. et al. Clandestinovirus: A Giant Virus With Chromatin Proteins and a Potential to Manipulate the Cell Cycle of Its Host *Vermamoeba vermiformis*. **Frontiers in Microbiology**, v. 12, 10 ago. 2021.
- SAHMI-BOUNSIAR, D. et al. Marseilleviruses: An Update in 2021. **Frontiers in Microbiology**, v. 12, p. 648731, 2 jun. 2021.
- SCHIMMEL, P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. **Nature Reviews Molecular Cell Biology** **2017 19:1**, v. 19, n. 1, p. 45–58, 6 set. 2017.
- SCHULZ, F. et al. Giant viruses with an expanded complement of translation system components. **Science**, v. 356, n. 6333, p. 82–85, 7 abr. 2017.
- SCHULZ, F.; ABERGEL, C.; WOYKE, T. Giant virus biology and diversity in the era of genome-resolved metagenomics. **Nature Reviews Microbiology** **2022 20:12**, v. 20, n. 12, p. 721–736, 28 jul. 2022.
- SHEPPARD, K. et al. From one amino acid to another: tRNA-dependent amino acid biosynthesis. **Nucleic Acids Research**, v. 36, n. 6, p. 1813–1825, 1 abr. 2008.
- SIGLER, P. B. **An analysis of the structure of tRNA**. [s.l.: s.n.]. Disponível em: <[www.annualreviews.org](http://www.annualreviews.org)>.
- SILVA, L. C. F. et al. Modulation of the expression of mimivirus-encoded translation-related genes in response to nutrient availability during *Acanthamoeba castellanii* infection. **Frontiers in Microbiology**, v. 6, n. JUN, p. 1, 1 jun. 2015.
- SIMON, D. et al. Phylogeny and Self-Splicing Ability of the Plastid tRNA-Leu Group I Intron. **Journal of Molecular Evolution**, v. 57, n. 6, p. 710–720, dez. 2003.
- SRINIVASAN, G.; JAMES, C. M.; KRZYCKI, J. A. Pyrrolysine encoded by UAG in archaea: Charging of a UAG-decoding specialized tRNA. **Science**, v. 296, n. 5572, p. 1459–1462, 24 maio 2002.
- STEWART, T. S.; ROBERTS, R. J.; STROMINGER, J. L. Novel Species of tRNA. **Nature** **1971 230:5288**, v. 230, n. 5288, p. 36–38, 1971.

STOTHARD, P.; GRANT, J. R.; VAN DOMSELAAR, G. Visualizing and comparing circular genomes using the CGView family of tools. **Briefings in Bioinformatics**, v. 20, n. 4, p. 1576, 27 mar. 2019.

SUGAHARA, J. et al. Disrupted tRNA gene diversity and possible evolutionary scenarios. **Journal of Molecular Evolution**, v. 69, n. 5, p. 497–504, 14 nov. 2009.

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. **Molecular Biology and Evolution**, v. 38, n. 7, p. 3022–3027, 1 jul. 2021.

VAN DER GULIK, P. T. S. et al. On distinguishing between canonical tRNA genes and tRNA gene fragments in prokaryotes. **RNA Biology**, v. 20, n. 1, p. 48–58, 2023.

WALSH, D.; MOHR, I. Viral subversion of the host protein synthesis machinery. **Nature Reviews. Microbiology**, v. 9, n. 12, p. 860, dez. 2011.

WANG, N. et al. Evolutionary characteristics and phylogeny of cotton chloroplast tRNAs. **Planta**, v. 254, n. 6, p. 1–18, 1 dez. 2021.

WEISMAN, R. A.; KORN, E. D. Phagocytosis of Latex Beads by *Acanthamoeba*. I. Biochemical Properties. **Biochemistry**, v. 6, n. 2, p. 485–497, 1 fev. 1967.

WIDMANN, J. et al. Stable tRNA-based phylogenies using only 76 nucleotides. **RNA**, v. 16, n. 8, p. 1469, ago. 2010.

WILSON, W. H.; VAN ETTEN, J. L.; ALLEN, M. J. The Phycodnaviridae: The Story of How Tiny Giants Rule the World. **Current topics in microbiology and immunology**, v. 328, p. 1, 2009.

XIAN, Y.; XIAO, C. Current capsid assembly models of icosahedral nucleocytoviricota viruses. Em: **Advances in Virus Research**. [s.l.] Academic Press Inc., 2020. v. 108p. 275–313.

YOSHIHISA, T. Handling tRNA introns, archaeal way and eukaryotic way. **Frontiers in Genetics**, v. 5, n. JUL, 2014.

YOSHIKAWA, G. et al. Medusavirus, a Novel Large DNA Virus Discovered from Hot Spring Water. **Journal of Virology**, v. 93, n. 8, p. 2130–2148, 15 abr. 2019.

## 9. ANEXOS

### Anexo 1 – Tabela dos genomas utilizados no estudo

Grupo do vírus	Número de acesso no GenBank	Nome do isolado	Tamanho do genoma	% GC do genoma	Nº tRNAs codificados	
1	Yasminevirus	UPSH01000001.1 / UPSH01000002.1	Yasminevirus saudimassiliense	2126343	40,2	70
2	Tupanvirus	NC_075039.1	Tupanvirus salinum	1516267	29,1	71
3	Tupanvirus	NC_075038.1	Tupanvirus altamarinense	1439508	29,4	67
4	Mimivirus	NC_023639.1	Mimivirus terra2	1168989	27,9	7
5	Mimivirus	JF801956.1	Acanthamoeba castellanii mamavirus	1191693	28	7
6	Mimivirus	KF493731.1	Hirudovirus	1181042	28	7
7	Mimivirus	KF959826.2	Samba virus	1181380	28	7
8	Mimivirus	KU761889.1	Mimivirus Bombay	1182200	27,9	7
9	Mimivirus	NC_014649.1	Mimivirus bradfordmassiliense	1181549	28	7
10	Mimivirus	KM982401.1	Acanthamoeba polyphaga mimivirus Oyster	1200220	28	7
11	Mimivirus	JX975216.1	Megavirus courdo11	1246126	25,3	6
12	Mimivirus	KU877344.1	Megavirus powaiense	1208707	25,3	5
13	Mimivirus	MG807319.1	Megavirus vitis	1242360	25,2	5
14	Mimivirus	JX885207.1	Megavirus LBA111	1230522	25,3	5
15	Mimivirus	LT717347.1	Mimivirus U306	1181032	27,5	5
16	Mimivirus	KM982402.1	Mimivirus lagoaense	1221932	27,6	5
17	Mimivirus	MG807320.1	Moumouvirus australiense	1098002	25,1	9
18	Mimivirus	MK978772.1	Moumouvirus maliensis	999513	25,3	5
19	Mimivirus	NC_075037.1	Moumouvirus goulettemassiliense	1016844	24,7	5
20	Mimivirus	MH046811.2	Megavirus baoshanense	1224839	25,1	9
21	Mimivirus	MW004169.1	Mimivirus reunion Queen	1196989	27,9	7
22	Mimivirus	MW685904.1	Acanthamoeba polyphaga mimivirus M4	1024623	27,9	4
23	Mimivirus	NC_016072.1	Megavirus chilense	1259197	25,2	4
24	Mimivirus	NC_020104.1	Moumouvirus moumou	1021348	24,6	3
25	Kaumoebavirus	NC_034249.1	Kaumoebavirus Sc	350731	43,7	0
26	Kaumoebavirus	MT334784.1	Kaumoebavirus strain KLCC10	362586	43,7	0
27	Pacmanvirus	MZ440852.1	Pacmanvirus S19	418588	33,2	1
28	Pacmanvirus	NC_034383.1	Pacmanvirus A23	395405	33,6	1
29	Faustovirus	KU702949.1	Faustovirus strain D5b	464523	37,7	1
30	Faustovirus	KU702951.1	Faustovirus strain D6	462011	37,7	1
31	Faustovirus	KU556803.1	Faustovirus strain D3	455803	37,8	1
32	Faustovirus	KJ614390.1	Faustovirus strain E12	466265	36,2	0
33	Faustovirus	MN956669.1	Faustovirus strain VV10	456714	37,7	0
34	Faustovirus	MT335755.1	Faustovirus strain E9	491024	39,6	0
35	Faustovirus	MN830298.1	Faustovirus strain VV63	479542	39,7	0
36	Faustovirus	MN830297.1	Faustovirus strain vv57	478172	39,7	0
37	Faustovirus	MN830296.1	Faustovirus strain S17	476423	39,6	0
38	Faustovirus	MN830295.1	Faustovirus strain M6	472803	39,8	0
39	Faustovirus	MN830294.1	Faustovirus strain LCD7	477407	39,9	0
40	Faustovirus	MN534311.1	Faustovirus strain liban	470731	36,7	0
41	Faustovirus	KU702952.1	Faustovirus strain E23	465956	36,2	0
42	Faustovirus	KU702950.1	Faustovirus strain D5a	466051	36,2	0
43	Faustovirus	KU702948.1	Faustovirus strain E24	466012	36,2	0
44	Faustovirus	MK506267.1	Faustovirus mariensis	466080	36,2	0
45	Faustovirus	LT839607.1	Faustovirus ST1	470659	36,7	0
46	Mollivirus	NC_027867.1	Mollivirus sibericum	651523	60,1	5
47	Mollivirus	MN812837.1	Mollivirus kamchatka	648864	60,1	6
48	Pandoravirus	LC625835.1	Pandoravirus japonicus	1787268	64	15
49	Pandoravirus	NC_022098.1	Pandoravirus salinum	2473870	61,7	17
50	Pandoravirus	NC_026440.1	Pandoravirus inopinatum	2243109	60,7	9
51	Pandoravirus	NC_037667.1	Pandoravirus quercus	2077288	60,7	11
52	Pandoravirus	MK174290.1	Pandoravirus celtis	2028440	60,8	9
53	Pandoravirus	NC_021858.1	Pandoravirus dulcis	1908524	63,7	18
54	Pandoravirus	NC_037665.1	Pandoravirus macleodensis	1838258	57,9	9
55	Pandoravirus	NC_037666.1	Pandoravirus neocaledonia	2003191	60,6	9
56	Medusavirus	AP018495.1	Medusavirus medusae	381277	61,7	6
57	Medusavirus	MW018138.1	Medusavirus sthenus	362811	62,6	1
58	Clandestinovirus	MZ420154.1 / MZ420155.1	Clandestinovirus	581987	43,5	1
59	Marseillevirus	NC_013756.1	Marseillevirus marseillevirus T19	368454	44,7	2
60	Marseillevirus	NC_025412.1	Melbournevirus	369360	44,7	1
61	Marseillevirus	NC_031465.1	Golden Marseillevirus	360610	43,1	2
62	Marseillevirus	MG827395.1	Marseillevirus Shanghai	368078	44,7	0
63	Marseillevirus	KY073338.1	Kurlavirus BKC-1	361368	42,9	0
64	Marseillevirus	NC_029692.1	Brazilian marseillevirus	362276	43,3	0
65	Marseillevirus	KF261120.1	Cannes 8 virus	374041	44,6	0
66	Marseillevirus	NC_033775.1	Noumeavirus	376207	42,9	0
67	Orpheovirus	NC_036594.1	Orpheovirus IHUMI-LCC2	1473573	25	2
68	Cedratvirus	MN873693.1	Cedratvirus kamchatka	466767	41,5	1
69	Cedratvirus	NC_032108.1	Cedratvirus A11	589068	40,6	1
70	Pithovirus	NC_023423.1	Pithovirus sibericum	610033	35,8	0

**Anexo 2 – Mapa do local e ano de isolamento dos vírus gigantes pelo globo**

