

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

**Instituto de Ciências Biológicas**

**Programa Interunidades de Pós-graduação em Bioinformática**

THIAGO MENDONÇA DOS SANTOS

**EXPLORANDO A DIVERSIDADE GENÉTICA E TAXONÔMICA NO GRUPO**

**EXTENDED *Asfarviridae*: Um estudo de genômica comparativa.**

Belo Horizonte

2024

THIAGO MENDONÇA DOS SANTOS

**EXPLORANDO A DIVERSIDADE GENÉTICA E TAXONÔMICA NO GRUPO  
EXTENDED *Asfarviridae*: Um estudo de genômica comparativa.**

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

Orientador: Dr. Luiz Eduardo Vieira Del-Bem.

Belo Horizonte

2024

043

Santos, Thiago Mendonça dos.

Explorando a diversidade genética e taxonômica no grupo extended Asfarviridae: uma análise de genômica comparativa [manuscrito] / Thiago Mendonça dos Santos. – 2024.

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

Orientador: Dr. Luiz Eduardo Vieira Del-Bem.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Asfarviridae. 3. Genômica. 4. Vírus Gigantes. 5. Taxonomia I. Del-Bem, Luiz Eduardo Vieira. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS

**FOLHA DE APROVAÇÃO**

**Thiago Mendonça dos Santos**

**"Explorando a diversidade genética e taxonômica no grupo extended Asfarviridae: Uma análise de genômica comparativa"**

Tese aprovada pela banca examinadora constituída pelos Professores:

Prof. Luiz Eduardo Vieira Del Bem - Orientador

UFMG

Prof. André Felipe Streck

UCS

Prof. Siomar de Castro Soares

UFTM

Prof. Rodrigo Araújo Lima Rodrigues

UFMG

Prof. Pedro Milet Meirelles

UFBA

Belo Horizonte, 19 de junho de 2024.



Documento assinado eletronicamente por **Rodrigo Araújo Lima Rodrigues, Professor do Magistério Superior**, em 20/06/2024, às 11:49, 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 **Siomar de Castro Soares, Usuário Externo**, em 20/06/2024, às 13:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Pedro Milet Meirelles, Usuário Externo**, em 21/06/2024, às 11:19, 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 **Andre Felipe Streck, Usuário Externo**, em 24/06/2024, às 15:57, 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 **Luiz Eduardo Vieira Del Bem, Professor do Magistério Superior**, em 26/06/2024, às 14:47, 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 **3296777** e o código CRC **EC956A13**.

---

## AGRADECIMENTOS

Primeiramente, agradeço ao meu orientador, Dr. Luiz Eduardo Vieira DelBem, pela orientação extraordinária e pelas valiosas sugestões. Sua didática, pensamento inovador e constante estímulo foram fundamentais não somente para o desenvolvimento deste trabalho, mas também para o meu desenvolvimento pessoal, político e profissional. O pesquisador que sou hoje é em grande parte atribuído a ele. Sou extremamente grato e me sinto muito sortudo pela oportunidade de o ter tido como orientador.

Ao Dr. Jônatas Abrahão, pelo auxílio na idealização deste projeto, pelas valiosas contribuições e por ser uma fonte de orgulho e inspiração na UFMG.

Aos meus colegas de pesquisa no DelBem Lab, especialmente Breno Lisboa e Wenderson Rodrigues, por estarem sempre solícitos quando precisei e por serem fontes de inspiração ao longo da minha jornada. A parceria e a amizade formadas durante esse período foram, também, de extrema importância. Agradeço imensamente pelo conhecimento adquirido e pelas valiosas lições compartilhadas ao longo do tempo, que foram, sem dúvida, fundamentais para meu crescimento acadêmico e profissional.

A Marcus, agradeço por ter sido minha base, meu suporte, em todos os sentidos, especialmente na reta final (e por ter me ajudado com algumas análises, figuras e slides). Seu encorajamento constante e seu apoio nos momentos de desânimo, foram fundamentais para me manter focado e determinado. Cada palavra de incentivo, cada gesto de confiança renovava minha determinação me lembrando do meu potencial. Sua presença foi mais do que um apoio; foi um alento. Obrigado!

Agradeço profundamente à minha família pelo apoio incondicional ao longo desta jornada. Sem o amor, encorajamento e suporte deles, nada disso seria possível. Em especial, expresso minha gratidão à minha avó, Jovelina, cuja dedicação incansável e inspiração constante foram fundamentais em cada passo do caminho. Nunca mediu esforços em me ajudar, movendo céus e terra. Sou imensamente grato por sua presença e dedicação que sempre me cercou.

Ao Programa de Pós-Graduação em Bioinformática pela alta qualidade.

A CAPES, pelo financiamento pessoal durante boa parte do Doutorado.

*“O verdadeiro início da atividade científica consiste mais em descrever fenômenos e depois em proceder para agrupá-los, classificá-los e correlacioná-los” (Freud S., 1915).*

## RESUMO

A família de vírus gigantes *Asfarviridae* abrange oficialmente apenas espécies do gênero *Asfivirus*, entretanto, observa-se uma associação genética com outros vírus, como Faustovírus, Kaumoebavírus, Pacmanvírus e *Abalone asfar-like virus*, formando assim o grupo *extended Asfarviridae*. Apesar da associação genética, há uma falta de consenso sobre a classificação desses vírus como um táxon coeso. Esse trabalho visa investigar a diversidade genômica e a taxonomia dos vírus do grupo *extended Asfarviridae*. A metodologia adotada abrangeu uma análise detalhada das características genômicas desses vírus, incluindo tamanho de genoma, quantidade de proteínas preditas e conteúdo G+C, identidades médias de aminoácido (AAI) e nucleotídeo (ANI), frequentemente utilizadas em taxonomia viral. Além disso, foi realizado um estudo do pangenoma do grupo *extended Asfarviridae* ao agrupar as sequências de proteínas em grupos de ortólogos. Adicionalmente, a natureza aberta ou fechada do genoma foi investigada, bem como a identificação de genes *core* e acessórios. A análise filogenética, realizada por meio de Máxima Verossimilhança, permitiu a construção de uma árvore filogenética com base nos genes *core*, proporcionando entendimentos sobre a relação evolutiva entre esses vírus. A investigação da manutenção da colinearidade gênica entre os diferentes grupos de vírus também foi realizada. Os resultados obtidos revelaram uma diversidade genômica considerável dentro do grupo *extended Asfarviridae*, o que contrasta com a uniformidade esperada em membros de um mesmo táxon. Essa diversidade se manifesta em diferenças significativas nos genomas, tanto em termos de tamanho quanto de conteúdo G+C. O pangenoma de *extended Asfarviridae* mostrou-se aberto, com um *core* restrito, de 37 genes, e um genoma acessório amplo. Além disso, foram observadas variações na identidade de aminoácidos e nucleotídeos entre os membros desse grupo, indicando uma ampla diversidade genética. Adicionalmente, foi identificada uma falta de conservação na ordem gênica entre esses vírus. Com base nos resultados obtidos, juntamente com os dados fenotípicos, propõe-se uma revisão na taxonomia desses vírus, que não mais os agrupe em uma única família. Sugere-se, portanto, a possível separação de Faustovírus e *Asfivirus*, Kaumoebavírus e Pacmanvírus em múltiplas espécies distintas. Essas descobertas enfatizam a necessidade contínua de investigação para uma compreensão mais precisa da diversidade genética e evolutiva dos vírus gigantes, além de suas implicações taxonômicas e biológicas.

Palavras-chave: *Asfarviridae*; genômica comparativa; vírus gigantes; pangenoma; taxonomia.



## ABSTRACT

The giant virus family *Asfarviridae* officially encompasses only species of the genus *Asfivirus*, however, a genetic association with other viruses such as Faustovirus, Kaumoebavirus, Pacmanvirus, and Abalone asfar-like virus has been observed, thus forming the “extended *Asfarviridae*” group. Despite the genetic association, there is a lack of consensus on classifying these viruses as a cohesive taxon. This work aims to investigate the genomic diversity and taxonomy of viruses in the extended *Asfarviridae* group. The adopted methodology included a detailed analysis of the genomic characteristics of these viruses, including genome size, predicted protein count, and G+C content, as well as average amino acid (AAI) and nucleotide (ANI) identities, commonly used in viral taxonomy. Additionally, a study of the extended *Asfarviridae* group's pangenome was conducted by clustering protein sequences into orthologous groups. Furthermore, the open or closed nature of the genome was investigated, along with the identification of core and accessory genes. Phylogenetic analysis, performed through Maximum Likelihood, allowed the construction of a phylogenetic tree based on core genes, providing insights into the evolutionary relationship between these viruses. The investigation of gene collinearity preservation among different virus groups was also conducted. The obtained results revealed considerable genomic diversity within the extended *Asfarviridae* group, contrasting with the expected uniformity among members of the same taxon. This diversity is manifested in significant genome differences, both in terms of size and G+C content. The extended *Asfarviridae* pangenome demonstrated to be open, with a restricted core of 37 genes and a broad accessory genome. Additionally, variations in amino acid and nucleotide identity were observed among members of this group, indicating wide genetic diversity. Furthermore, a lack of conservation in gene order among these viruses was identified. Based on the obtained results, together with phenotypic data, a revision in the taxonomy of these viruses is proposed, no longer grouping them into a single family. Therefore, the possible separation of Faustovirus and *Asfivirus*, Kaumoebavirus and Pacmanvirus into multiple distinct species is suggested. These findings emphasize the ongoing need for research for a more precise understanding of giant virus genetic and evolutionary diversity, as well as their taxonomic and biological implications.

Keywords: *Asfarviridae*; comparative genomics; giant viruses; pangenoma; taxonomy.

## LISTA DE ILUSTRAÇÕES

- Figura 1** – Média dos tamanhos de partícula e de genoma de diferentes famílias/grupos de vírus gigantes.....19
- Figura 2** - Classificação hierárquica utilizada na taxonomia de vírus associada às metodologias que podem ser empregadas para determinar as relações evolutivas. AAS - Similaridade de Sequência de Aminoácidos; NS - Similaridade de Sequência de Nucleotídeos.....21
- Figura 3** – Árvore filogenética baseada em sete genes marcadores compartilhados entre todos os membros do filo *Nucleocyotviricota* (NCLDVs). Ordens virais representadas por diferentes cores.....25
- Figura 4** – Arvore filogenética de *extended Afarviridae* baseada no gene da DNA polimerase B. A raiz da árvore está baseada em vírus de *Poxviridae*.....29
- Figura 5** – Estrutura do vírion do ASFV. (A) Micrografia eletrônica do vírion, barra representa 200 nm (B) Reconstrução cryo-EM do capsídeo do ASFV cepa BA71V, colorida de acordo com a distância radial do centro do vírus (C) Representação esquemática e simplificada da estrutura multicamada do vírion do ASFV.....30
- Figura 6** – Hospedeiros conhecidos do ASFV. (A) Carrapato (*Ornithodoros* spp.), vetor. (B) Javali Africano (*Phacochoerus africanus*), (C) Porco doméstico (*Sus domesticus*), (D) Javali (*Sus scrofa*).....32
- Figura 7** – Faustovírus e seu hospedeiro conhecido. (A) *Vermamoeba vermiformis*, protozoário hospedeiro dos faustovírus. (B) Micrografia eletrônica do vírion do Faustovirus E12, (C) Reconstrução cryo-EM do capsídeo, colorida de acordo com a distância radial do centro do vírus, (D) Representação esquemática e simplificada da estrutura multicamada do vírion.....35
- Figura 8** – Micrografia eletrônica do vírion do Kaumobavirus KLCC10.....37
- Figura 9** – Estrutura do Pacmanvirus e seu hospedeiro conhecido. (A) *Acanthamoeba castellanii*, hospedeiro do Pacmanvirus, (B) Micrografia eletrônica do vírion do Pacmanvirus A23, (C) Reconstrução cryo-EM do capsídeo, colorida de acordo com a distância radial do centro do vírus, (D) Representação esquemática e simplificada da estrutura multicamada do vírion.....39
- Figura 10** – AbALV e seu hospedeiro conhecido. (A) Micrografia eletrônica do vírion do AbALV. (B) *Haliotis madaka*, uma das espécies de abalone que servem como hospedeiro para AbALV.....40
- Figura 11** - Diagrama de Venn ilustrando os subgrupos de um pangenoma formado por três organismos.....42

## LISTA DE ILUSTRAÇÕES

- Figura 12** – *Boxplots* representando a variação em tamanho do genoma, número de proteínas preditas e conteúdo G+C em 39 membros de *extended Asfarviridae* separados por grupo. Letras minúsculas diferentes representam diferença estatisticamente significante (Mann-Whitney  $p < 0.05$ ).....52
- Figura 13** – *Flower Plot* mostrando os genes *core*, acessórios e *singletons* de 39 membros de *extended Asfarviridae*. O *flower Plot* exhibe o número de genes *core* (no centro), o número de genes acessórios (dentro do anel) e o número de genes *singletons* (nas pétalas). AbALV – *Abalone arfar-like virus*; ASFV – *African swine fever virus*.....53
- Figura 14** – *Heatmap* representando a proporção de genes compartilhados entre cada par de membros de *extended Asfarviridae*. AbALV – *Abalone arfar-like virus*; ASFV – *African swine fever virus*.....54
- Figura 15** – Curva de acumulação de genes para o pangenoma total e genes *core* de *extended Asfarviridae*. O valor de  $\alpha$  indica a lei de potência de Heaps para o número médio de novos genes por genoma após 1.000 permutações do pangenoma. Abaixo de 1 indica que o pangenoma está em estado "aberto" (Tettelin et al., 2005).....55
- Figura 16** – Histograma das distribuições dos genes nos grupos pangenômicos *core*, *soft core*, *shell* e *cloud*. O gráfico de pizza exhibe o número total de genes com as porcentagens entre parênteses.....56
- Figura 17** – Tamanho médio das proteínas preditas do pangenoma de *extended Asfarviridae* separadas por grupo pangenômico ou total. AbALV – *Abalone asfar-like virus*. Letras minúsculas diferentes representam diferença estatisticamente significante (Mann-Whitney  $p < 0.05$ ).....57
- Figura 18**– Análise de Componentes Principais (PCA) baseada na presença/ausência de genes acessórios para 39 genomas de *extended Asfarviridae*. Os indivíduos são plotados de acordo com seus escores nos dois maiores componentes principais. As variâncias explicadas para cada componente principal se encontram entre parênteses.....58
- Figura 19** – Classificação funcional dos clusters de proteínas ortólogas do pangenoma de *extended Asfarviridae* com base no banco de dados do *Clusters of Orthologous Groups* (COG). Todas as proteínas preditas foram alinhadas ao banco de dados COG a fim de prever funções. As letras maiúsculas no eixo Y indicam as categorias COG conforme legenda. A porcentagem sinalizada abaixo do grupo pangenômico indica a proporção de transcritos classificados para a totalidade de cada grupo.....59
- Figura 20** – Alinhamento de estruturas tridimensionais de proteínas hipotéticas presentes no genoma *core* de *extended Asfarviridae* com estruturas homólogas. RMSD – Root Mean Square Deviation.....63

## LISTA DE ILUSTRAÇÕES

- Figura 21** – Alinhamento de múltiplas sequências de amino ácidos do gene *Major Capsid Protein* (MCP) presente em 39 genomas de *extended Asfarviridae*. Genomas de faustovírus possuem 2,6 sequencias por genoma em média.....64
- Figura 22** – Árvore filogenética de Máxima Verossimilhança com base em 37 genes *core* concatenados de *extended Asfarviridae*. Valores de suporte de ramo são baseados em SH-aLRT.....65
- Figura 23** – *Heatmap* de valores de identidade média de aminoácidos (AAI) par-a-par entre 39 genomas de *extended Asfarviridae*. AbALV – *Abalone asfar-like virus*, ASFV – *African Swine Fever Virus*.....66
- Figura 24** – *Heatmap* de valores de identidade média de aminoácidos (AAI) par-a-par entre todos os genomas agrupados em seus respectivos grupos. AbALV – *Abalone asfar-like virus*; ASFV – *African Swine Fever Virus*.....68
- Figura 25** – Médias de valores par-a-par de identidade de aminoácidos (AAI) entre membros de cada família de NCLDV. Barras de erro indicam desvio padrão da média..68
- Figura 26** – *Heatmap* de valores de identidade média de nucleotídeos (ANI) par-a-par entre (A) 18 genomas de *Asfivirus* e (B) 16 genomas de Faustovírus. ASFV – *African Swine Fever Virus*.....70
- Figura 27** – Análise de sintenia de 39 membros da família *extended Asfarviridae* (Linhas azuis: identidade mínima de 70%, Linhas vermelhas: identidade mínima de 20%). Cada bloco correspondente representa a comparação de colinearidade de dois vírus. Se não houver sequência de aminoácidos colinear entre dois vírus, o bloco estará em branco. AbALV – *Abalone asfar-like virus*; ASFV – *African Swine Fever Virus*.....72

## LISTA DE TABELAS

**Tabela 1.** Lista de genomas de *extended Asfarviridae* selecionados para o estudo. Descritos estão o número de proteínas preditas na anotação, o comprimento do genoma em *base pairs* (bp), a arquitetura do genoma, o conteúdo G+C e o identificador do NCBI.....50

**Tabela 2.** Genes do grupo *core* do pangenoma de *extended Asfarviridae* e suas funções preditas por diferentes metodologias. Marcados em azul claro estão os únicos genes que não são encontrados no genoma *core* de NCLDV.....60

## LISTA DE ABREVIATURAS E SIGLAS

AAI	<i>Average Aminoacid Identity</i> (Identidade Média de Aminoácidos)
AbALV	<i>Abalone asfar-like virus</i>
ANI	<i>Average Nucleotide Identity</i> (Identidade Média de Nucleotídeos)
APMV	<i>Acanthamoeba polyphaga Mimivirus</i>
ASFV	<i>African Swine Fever Virus</i> (Vírus da Peste Suína Africana)
BLAST	<i>Basic Local Alignment Search Tool</i> (Ferramenta básica de pesquisa de alinhamento local)
COG	<i>Cluster of Orthologous Groups</i> (Cluster de grupos ortólogos)
dsDNA	DNA de fita dupla
dsRNA	RNA de fita dupla
HGT	<i>Horizontal Gene Transfer</i> (Transferência Gênica Horizontal)
ICTV	<i>International Committee on Taxonomy of Viruses</i> (Comitê Internacional de Taxonomia de Vírus)
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i> (Enciclopédia de Genes e Genomas de Kyoto)
MCP	<i>Major Capsid Protein</i> (Proteína do Capsídeo)
NCBI	<i>National Center for Biotechnology Information</i> (Centro Nacional de informações sobre biotecnologia)
NCLDV <sub>s</sub>	<i>Nucleo-Cytoplasmic Large DNA Viruses</i> (Vírus de DNA grande nucleocitoplasmático)
nr	<i>Non-Redundant Database</i> (Banco de dados do NCBI)
ORFs	<i>Open Reading Frames</i> (Quadros de Leitura Aberta)
PCA	<i>Principal Components Analysis</i> (Análise de Componentes Principais)
RMSD	<i>Root Mean Square Deviation</i> (Raiz quadrada do erro-médio)
ssDNA	DNA de fita única
ssRNA	RNA de fita única

## SUMÁRIO

Página

<b>1. INTRODUÇÃO E JUSTIFICATIVA.....</b>	<b>15</b>
<b>2. REVISÃO BIBLIOGRÁFICA.....</b>	<b>17</b>
2.1 <i>VÍRUS GIGANTES</i> .....	17
2.2 <i>TAXONOMIA VIRAL</i> .....	20
2.3 <i>FILOGENIA E EVOLUÇÃO DE VÍRUS GIGANTES</i> .....	23
2.4 <i>EXTENDED Asfarviridae</i> .....	26
2.4.1 <i>Asfivirus</i> .....	28
2.4.2 <i>Faustovirus</i> .....	32
2.4.3 <i>Kaumoebavirus</i> .....	34
2.4.4 <i>Pacmanvirus</i> .....	36
2.4.5 <i>Abalone asfar-like virus (AbALV)</i> .....	38
2.5 <i>CONCEITOS FUNDAMENTAIS DE PANGENOMA</i> .....	39
<b>3. OBJETIVOS.....</b>	<b>44</b>
3.1 <i>OBJETIVO GERAL</i> .....	44
3.2 <i>OBJETIVOS ESPECÍFICOS</i> .....	44
<b>4. MATERIAL E MÉTODOS.....</b>	<b>45</b>
4.1 <i>OBTENÇÃO DOS GENOMAS E ANOTAÇÃO GÊNICA</i> .....	45
4.2 <i>AGRUPAMENTO DAS SEQUÊNCIAS PROTEICAS</i> .....	45
4.3 <i>DESCRIÇÃO DAS CARACTERÍSTICAS GENÔMICAS</i> .....	46
4.4 <i>ANÁLISE FILOGENÉTICA COM BASE NOS GENES CORE</i> .....	46
4.5 <i>ANÁLISE FUNCIONAL DAS PROTEÍNAS</i> .....	47
4.6 <i>CÁLCULO DE AAI e ANI</i> .....	47
4.7 <i>ANÁLISE DE SINTENIA</i> .....	48
<b>5. RESULTADOS.....</b>	<b>49</b>
5.1 <i>VARIAÇÕES GENÔMICAS SIGNIFICATIVAS EM EXTENDED Asfarviridae</i> .....	49
5.2 <i>EXTENDED Asfarviridae APRESENTA UM PANGENOMA ABERTO EXTENSO E UM CORE LIMITADO</i> .....	51
5.3 <i>ANOTAÇÃO FUNCIONAL EM EXTENDED Asfarviridae APRESENTA DESAFIOS</i> .....	58
5.4 <i>OS GRUPOS DE extended Asfarviridae APRESENTAM ASSINATURAS PARTICULARES DE AMINOÁCIDOS E NUCLEOTÍDEOS</i> .....	65
5.4 <i>ANÁLISE DE SINTENIA REVELA VARIAÇÕES NA CONSERVAÇÃO DA ORDEM GÊNICA EM VÍRUS DE EXTENDED Asfarviridae</i> .....	70
<b>6. DISCUSSÃO.....</b>	<b>72</b>
<b>7. CONCLUSÃO.....</b>	<b>81</b>
<b>8. PERSPECTIVAS.....</b>	<b>83</b>
<b>REFERÊNCIAS.....</b>	<b>84</b>
<b>ANEXO A - PRODUÇÃO ACADÊMICA.....</b>	<b>108</b>

## 1. INTRODUÇÃO E JUSTIFICATIVA

A definição abrangente e precisa de vírus permanece desafiadora, dada sua natureza dinâmica e diversificada (Koonin et al., 2021). Inicialmente conceituados por Lwoff na década de 1950, os vírus foram caracterizados como elementos genéticos simples e móveis, dependentes de um hospedeiro celular para replicação e envolvidos por uma capa proteica (Lwoff, 1957; Lwoff & Tournier, 1966). Contudo, o paradigma viral foi profundamente alterado em 2003 com a descoberta do *Acanthamoeba polyphaga Mimivirus* (APMV), desafiando as características tradicionais dos vírus (Scola et al., 2003).

Os vírus gigantes, exemplificados pelo APMV, apresentam características singulares, como vírions visíveis ao microscópio de luz, genomas maiores que alguns organismos unicelulares, repertório gênico inédito e diversificado (Raoult et al., 2004). Essa descoberta revolucionou a virologia, questionando conceitos estabelecidos, destacando a necessidade de revisitar as definições convencionais e criando a necessidade de novos estudos de isolamento e sequenciamento a fim de se descobrir novos vírus desse tipo, por muito tempo não estudados. Este cenário motivou a formação e uma possível expansão da família de vírus gigantes *Asfarviridae*, originando o clado "*extended Asfarviridae*" para incluir *Asfivirus* e vírus geneticamente relacionados, como faustovírus, kaumobavírus, pacmanvírus e *Abalone asfar-like virus* (AbALV) (Andreani et al., 2017).

Contudo, persiste uma lacuna de entendimento entre os pesquisadores quanto à inclusão oficial desses vírus associados na família *Asfarviridae*. Surgem questionamentos acerca de sua possível condição como membros fundadores de outras famílias, além da busca pela elucidação das reais associações genéticas e evolutivas entre essas entidades virais. Essa incerteza evidencia a complexidade subjacente à classificação taxonômica desses vírus, suscitando a necessidade contínua de estudos aprofundados para clarificar essas relações e consolidar a compreensão da diversidade e evolução no contexto desse clado.

Neste contexto, a realização de uma caracterização genômica e funcional abrangente, juntamente com a construção de um pangenoma que reflita o conteúdo gênico de *extended Asfarviridae*, desempenha um papel crucial na ampliação do nosso entendimento sobre a diversidade viral desse grupo. Isso nos capacita a responder a questões fundamentais sobre sua classificação como uma família e proporcionar uma estrutura organizada que reflete precisamente as relações evolutivas entre esses organismos. Essa contribuição é alcançada ao oferecer esclarecimentos aprofundados sobre padrões de variabilidade genética, identificação



de marcadores genéticos distintivos, eventos evolutivos específicos e relações filogenéticas entre essas linhagens virais.

Para atingir esses objetivos, o presente estudo emprega uma abordagem *in silico*, combinando genomas disponibilizados publicamente de diversas espécies e isolados virais dentro desse clado, por meio de técnicas avançadas de bioinformática. Esse processo envolveu métodos como clusterização de proteínas por homologia, determinação de co-ortologia, análises de sequências de aminoácidos e nucleotídeos, bem como a ordem gênica. Essa investigação vai além da classificação taxonômica, explorando as similaridades e diferenças genéticas entre as espécies desse clado. Este enfoque mais amplo fornece dados cruciais para compreender a relação evolutiva entre as diferentes linhagens, desvendar a complexa estrutura genômica que as caracteriza e explorar a diversidade funcional que contribui para a ecologia viral única dentro desse clado. A análise da dinâmica de ganho e perda de novos genes dentro das diferentes espécies do clado também é possível. Esses eventos genéticos estão intrinsecamente associados à funcionalidade gênica dos vírus e à sua capacidade de se adaptar ao ambiente circundante. Compreender esses processos de adaptação é essencial para desvendar a complexidade da ecologia viral e as estratégias evolutivas adotadas por *extended Asfarviridae*.

Portanto, ao abordar essa lacuna no conhecimento, esse estudo não apenas contribuirá para a classificação mais precisa desses vírus como uma família, mas também lançará luz sobre os intrincados mecanismos genéticos subjacentes à sua diversidade, evolução e adaptação ambiental.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1 VÍRUS GIGANTES

Devido à sua natureza dinâmica e diversificada, atualmente ainda não existe uma definição precisa de vírus estabelecida (Koonin et al., 2021). Embora imprecisa, a definição corrente mais habitual, cunhada por Lwoff na década de 1950 baseado em critérios negativos, caracteriza os vírus como pequenos elementos genéticos móveis, simples que dependem de um hospedeiro celular para se multiplicar e são constituídos pelo seu material genético – seja RNA ou DNA – envolto por uma capa proteica protetora (Lwoff, 1957; Lwoff & Tournier, 1966).

Uma partícula viral completa, conhecida como vírion, desempenha o papel de entregar sua carga genética à célula hospedeira, o que viabiliza a expressão gênica por meio de transcrição e tradução, utilizando a maquinaria bioquímica das células eucarióticas ou procarióticas. Fora do ambiente celular hospedeiro, os vírus permanecem inertes por não possuírem por completo a maquinaria própria necessária para sua replicação (Gelderblom, 1996; Modrow et al., 2013).

O material genético viral, que pode ser RNA de fita única (ssRNA) e dupla (dsRNA) ou DNA de fita única (ssDNA) e dupla (dsDNA), frequentemente acompanhado por proteínas básicas, forma o nucleocapsídeo – um complexo que compreende o ácido nucleico associado a proteínas. Nos vírus envelopados, o nucleocapsídeo ainda é envolto por uma bicamada lipídica derivada da membrana da célula hospedeira. Devido ao seu tamanho físico reduzido, os genomas virais geralmente codificam um número limitado de proteínas estruturais e proteínas reguladoras não estruturais envolvidas no processo de replicação viral (Gelderblom, 1996; Modrow et al., 2013).

Posto isso, algumas das características presentes na clássica definição de vírus foram desafiadas no ano de 2003, quando um organismo que por onze anos havia sido classificado como bactéria passava a ser o primeiro vírus gigante de ameba descrito, o APMV (Scola et al., 2003). Apesar de compartilhar muitas características com os vírus canônicos, o APMV apresenta diversas características presentes em outros microrganismos. A maioria dos vírus é menor do que o menor comprimento de onda de luz, o que significa que não podem ser observados com microscopia de luz visível (Daaboul et al., 2017). Para o APMV e outros vírus gigantes, seus vírions, ao contrário, são visíveis através do microscópio de luz (Scola et al., 2003), têm genomas maiores que algumas bactérias e repertório gênico mais diverso que outros vírus (Raoult et al., 2004). Ainda, o APMV possui proteínas nunca antes encontradas em vírus como fatores de tradução e sintetases de aminoacil-tRNA além de seis tRNAs (Raoult et al., 2004). Além disso, ao contrário dos outros vírus conhecidos até então, o APMV pode entrar em

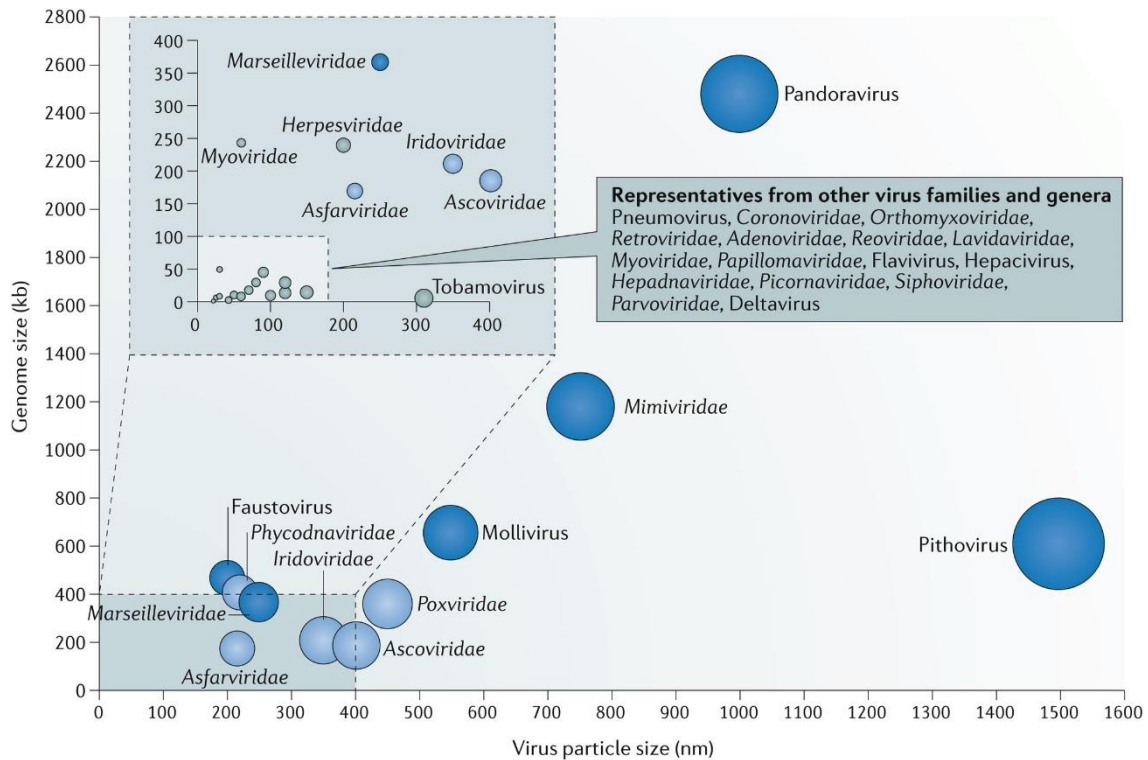
células hospedeiras por fagocitose devido ao seu grande tamanho, sem a necessidade de interações específicas vírus-hospedeiro, (Ghigo et al., 2008).

Vírus gigantes são vírus que apresentam uma ampla gama de formas e tamanhos de vírions. A maioria dos vírions possui tamanhos entre 200 e 400 nm (Schulz et al., 2022). O maior vírus descrito até o momento, em termos de tamanho físico, é o *Pithovirus sibericum*, com 1,500 nm de comprimento (Legendre et al., 2014), superando algumas bactérias pequenas, como *Mycoplasmas*, e até mesmo alguns Eucariotos unicelulares, a exemplo de algas da classe *Prasinophyceae*, as quais também são infectadas por outros vírus gigantes (Bellec et al., 2009; Sandaa et al., 2001). O *P. sibericum* foi descoberto em 2014 a partir de uma amostra de *permafrost* siberiano de mais de 30.000 anos (Legendre et al., 2014). Foi “revivido” e ainda demonstrou a capacidade de infectar amebas em condições laboratoriais. Esse resultado suscita discussões mais amplas, não apenas sobre a diversidade viral, mas também sobre os impactos do aquecimento global (Alempic et al., 2023; Khan et al., 2020).

Os vírus gigantes foram inicialmente ignorados devido às práticas de filtragem que presumiam que todos os vírus eram de tamanho diminuto. A descrição de vírus gigantes desafia as convenções tradicionais na virologia, especialmente a noção histórica de “filtrabilidade” como critério distintivo que separaria os vírus de outros microrganismos. Este conceito, que remonta ao isolamento do vírus da doença do mosaico do tabaco em 1892, presumia que todos os vírus seriam suficientemente pequenos para passarem através de filtros com poros de até 200 nm (Claverie & Abergel, 2016).

Além de seu tamanho físico, o genoma desses vírus é maior do que os outros vírus conhecidos e até mesmo maior que alguns genomas bacterianos, apresentando genomas acima de 2 Mb (**figura 1**), como no caso do *Pandoravirus salinus* (Philippe et al., 2013). Esses vírus possuem genomas de dsDNA, semelhantes aos de seus hospedeiros celulares. A variação no tamanho dos genomas desses vírus é significativa, oscilando de aproximadamente 100 kb em alguns iridovírus a cerca de 2,5 Mb em pandoravírus (Koonin & Yutin, 2019).

**Figura 1** – Média dos tamanhos de partícula e de genoma de diferentes famílias/grupos de vírus gigantes.



Fonte: Colson et al., 2017.

Ao longo dos anos, outras espécies de vírus gigantes foram descritas e agrupadas atualmente em onze famílias virais oficiais, todas pertencentes ao grupo taxonômico denominado *Nucleo-Cytoplasmic Large DNA Viruses* (NCLDVs), devido à característica de replicação no núcleo e citoplasma (Mönttinen et al., 2021), o qual hoje forma o filo *Nucleocyotoviricota* (Mönttinen et al., 2021). O avanço na descoberta desse grupo de vírus ocorreu através do seu isolamento em co-cultivo com eucariotos unicelulares, como amebas, particularmente *Acanthamoeba castellanii*, que serviu como um excelente sistema modelo. *A. castellanii* levou ao isolamento de diversas espécies de pitovírus, cedratvírus, faustovírus, pacmanvírus, marseillevírus, dentre diversos outros (Schulz et al., 2022). Além de *A. castellanii*, o uso de hospedeiros eucariotos unicelulares alternativos, como *Vermamoeba* spp. e espécies de microalgas, permitiu a identificação de várias outras espécies de vírus gigantes (Schulz et al., 2022).

Hoje, sabe-se que os vírus pertencentes a este filo infectam uma ampla variedade de hospedeiros eucarióticos, que vão desde os menores coanoflagelados unicelulares conhecidos até animais e algas multicelulares (Ruiz Martínez et al., 2023; Sun et al., 2020). Vírus gigantes que infectam *Acanthamoeba* penetram na célula por fagocitose ou fusão de membranas (Sobhy, 2017). No caso de fagocitose, após a abertura do capsídeo, a membrana que reveste o interior da partícula viral funde-se com a membrana do fagossomo liberando o DNA do vírus no interior

celular juntamente com proteínas associadas no citoplasma (Schulz et al., 2022; Sobhy, 2017). A depender do tipo de vírus, a replicação segue caminhos distintos. Vírus como os mimivírus e pitovírus contam com sua própria maquinaria de transcrição dentro dos vírions junto com seus genomas. Dessa forma, permanecem no citoplasma, onde desenvolvem grandes fábricas virais sem a necessidade de adentrar o núcleo celular (Schulz et al., 2022). Em contraste, os pandoravírus e molivírus não têm maquinaria de transcrição nas partículas virais, portanto, precisam transportar seus genomas para o núcleo celular a fim de acessar as funções necessárias para iniciar seu ciclo de replicação (Schulz et al., 2022). Os marseillevírus, por sua vez, comportam-se como um intermediário entre vírus citoplasmáticos e nucleares. Apesar da ausência de maquinaria de transcrição nas partículas virais, ao recrutar a maquinaria necessária do núcleo para o citoplasma por determinado tempo, eles acabam instalando suas fábricas virais no citoplasma (Rodrigues et al., 2020). Com exceção dos phaeovírus, o ciclo de replicação dos vírus gigantes é lítico, culminando na montagem de vírions maduros que são liberados por exocitose ou lise celular (Müller et al., 1996; Ruiz Martínez et al., 2023).

## 2.2 TAXONOMIA VIRAL

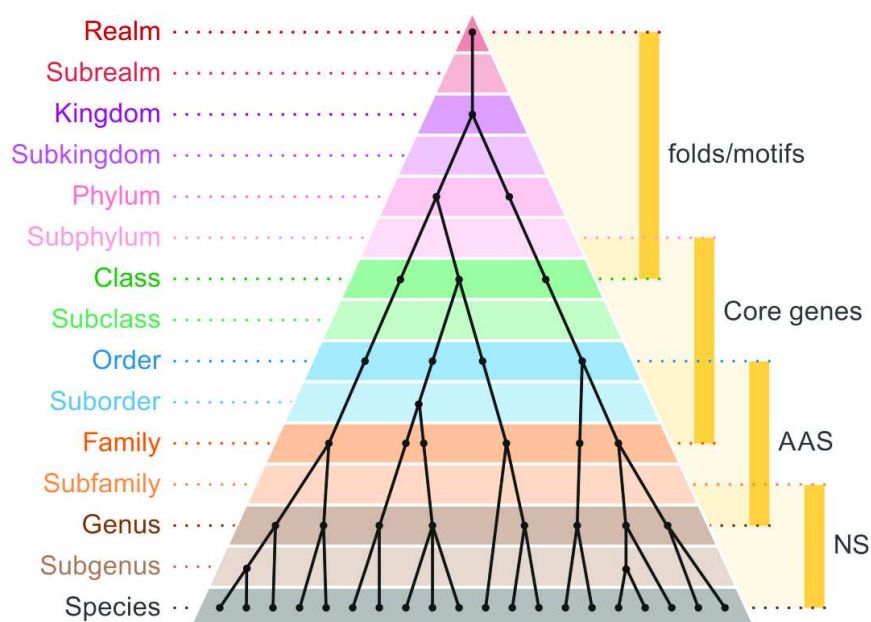
A taxonomia é uma disciplina fundamental no estudo de qualquer nova entidade biológica. Esse processo, que remonta pelo menos ao século XVIII, quando Carl Linnaeus definiu os princípios da taxonomia moderna, envolve a classificação e nomeação de organismos com base em características distintivas (Lefkowitz et al., 2018). A taxonomia é conduzida por cientistas que publicam seus trabalhos, apresentando evidências para a classificação proposta. À medida que novos dados surgem, seja sobre organismos previamente estudados ou relacionados, a hierarquia de classificação pode ser ajustada.

A taxonomia viral, por sua vez, fundamenta-se na análise e comparação de diversas características que delineiam a natureza do vírus. Essas características englobam a composição molecular (DNA ou RNA) e a configuração do genoma (simples ou dupla fita, linear ou circular, inteiro ou segmentado), a arquitetura e composição do capsídeo, as vias responsáveis pela expressão gênica viral, o local de replicação do material genético, a amplitude de hospedeiros, o modo de replicação, a patogenicidade e a similaridade de sequência (Adams et al., 2015; Lefkowitz et al., 2018).

Embora todas essas características desempenhem papel crucial, as comparações genômicas, tanto em termos de similaridade composicional quanto de filogenia, emergiram como elementos-chave na delimitação e distinção das categorias virais (Simmonds et al., 2017). Os métodos genômicos desempenham um papel crucial na delimitação das categorias virais,

principalmente em táxons mais basais, utilizando tanto análises baseadas em aminoácidos (AAS) quanto em nucleotídeos (NS) (**figura 2**). O alinhamento de sequências de aminoácidos e a análise de *Average Aminoacid Identity* (AAI), ou identidade média de aminoácidos, fornecem perspectivas sobre a composição e relacionamentos genéticos das proteínas virais, enquanto o alinhamento de sequências de nucleotídeos, conteúdo G+C e a *Average Nucleotide Identity* (ANI), ou identidade média de nucleotídeos, revelam similaridades no nível do DNA (Romiguier et al., 2010; Simmonds et al., 2023). Além disso, o suporte de ramificação (*branch support*) em filogenias fornece medidas estatísticas sobre a força de um clado filogenético, enquanto a sintenia de genes destaca a conservação estrutural genômica ao longo do tempo (Simmonds et al., 2017; Y. Wang et al., 2012). Esses métodos genômicos são essenciais para uma taxonomia viral precisa, proporcionando uma abordagem robusta e atualizada.

**Figura 2** - Classificação hierárquica utilizada na taxonomia de vírus associada às metodologias que podem ser empregadas para determinar as relações evolutivas. AAS - Similaridade de Sequência de Aminoácidos; NS - Similaridade de Sequência de Nucleotídeos.



Fonte: Simmonds et al., 2023.

Com base na avaliação dessas características, estabelece-se uma relação hierárquica que agrupa vírus com propriedades semelhantes. Essas propriedades definem táxons em diferentes níveis, desde reinos até espécies. Por exemplo, a ordem *Picornavirales* inclui vírus não envelopados, com partículas icosaédricas contendo um ou dois segmentos de RNA de sentido positivo. Membros da família *Picornaviridae*, uma das várias famílias pertencentes à ordem *Picornavirales*, apresentam um genoma não-segmentado e organização genômica conservada.

O gênero *Enterovirus* compartilha mais de 42% de identidade de aminoácidos. Na ponta final, exemplares da espécie *Enterovirus C* compartilham mais de 70% de identidade de aminoácidos, além de uma limitada gama de hospedeiros e receptores (Lefkowitz et al., 2018).

O atual relatório do Comitê Internacional de Taxonomia de Vírus (ICTV) lista 11,273 espécies de vírus e viroides distribuídas em 2,818 gêneros, 128 subfamílias, 264 famílias e 72 ordens (*Current ICTV Taxonomy Release* | ICTV, 2022). Há também um capítulo de vírus não classificados, representando possíveis novos gêneros ou famílias. A classificação taxonômica de vírus é uma ciência dinâmica, sujeita a atualizações regulares para refletir as últimas decisões taxonômicas. O ICTV é uma entidade internacional dedicada à taxonomia de vírus, incumbida de desenvolver, curar e manter uma taxonomia universal de vírus. Atuando com o apoio da divisão de virologia da União Internacional de Sociedades Microbiológicas, o ICTV desempenha um papel crucial na orientação das diretrizes para a nomeação de táxons e na aprovação da taxonomia e nomes propostos por pesquisadores antes de sua oficialização (*Introduction to Virus Taxonomy* | ICTV, 2011). O capítulo dedicado a vírus não classificados no nono relatório destaca a conscientização do ICTV sobre a presença de vírus que ainda não puderam, devido a características únicas, ser totalmente categorizados. Esses vírus representam potencialmente novos gêneros ou famílias, destacando a necessidade contínua de atualizações e revisões na classificação taxonômica para acomodar a crescente complexidade do mundo viral, principalmente advindo de estudos genômicos (Santos-Medellin et al., 2021; Suttle, 2005). Dentre os vírus sem classificação, encontram-se os faustovírus, kaumoebavírus, pacmanvírus e AbALV.

A diversidade viral é vasta, com estimativas sugerindo a existência de  $10^{31}$  partículas virais na Terra (Mushegian, 2020). As recentes técnicas de sequenciamento, como a metagenômica, têm revelado um viroma extenso e diversificado. No entanto, a classificação viral a partir desses dados desafia os critérios tradicionais, principalmente devido à falta de informações fenotípicas sobre muitos desses vírus detectados. Nesse caso, ainda se depende muito de abordagens tradicionais como isolamento desses vírus em culturas laboratoriais. A taxonomia viral não é uma ciência estática, mas sim um campo em constante evolução. À medida que avançamos na era da genômica, a classificação automática com base em sequências genômicas pode se tornar uma realidade.

Sua importância reside não apenas na organização do vasto mundo viral, mas também na compreensão das relações evolutivas entre esses agentes infecciosos, um desafio contínuo enfrentado pelos cientistas na busca por uma compreensão abrangente da vida microscópica. A correta classificação dos vírus é fundamental para compreender sua biologia, evolução e, mais

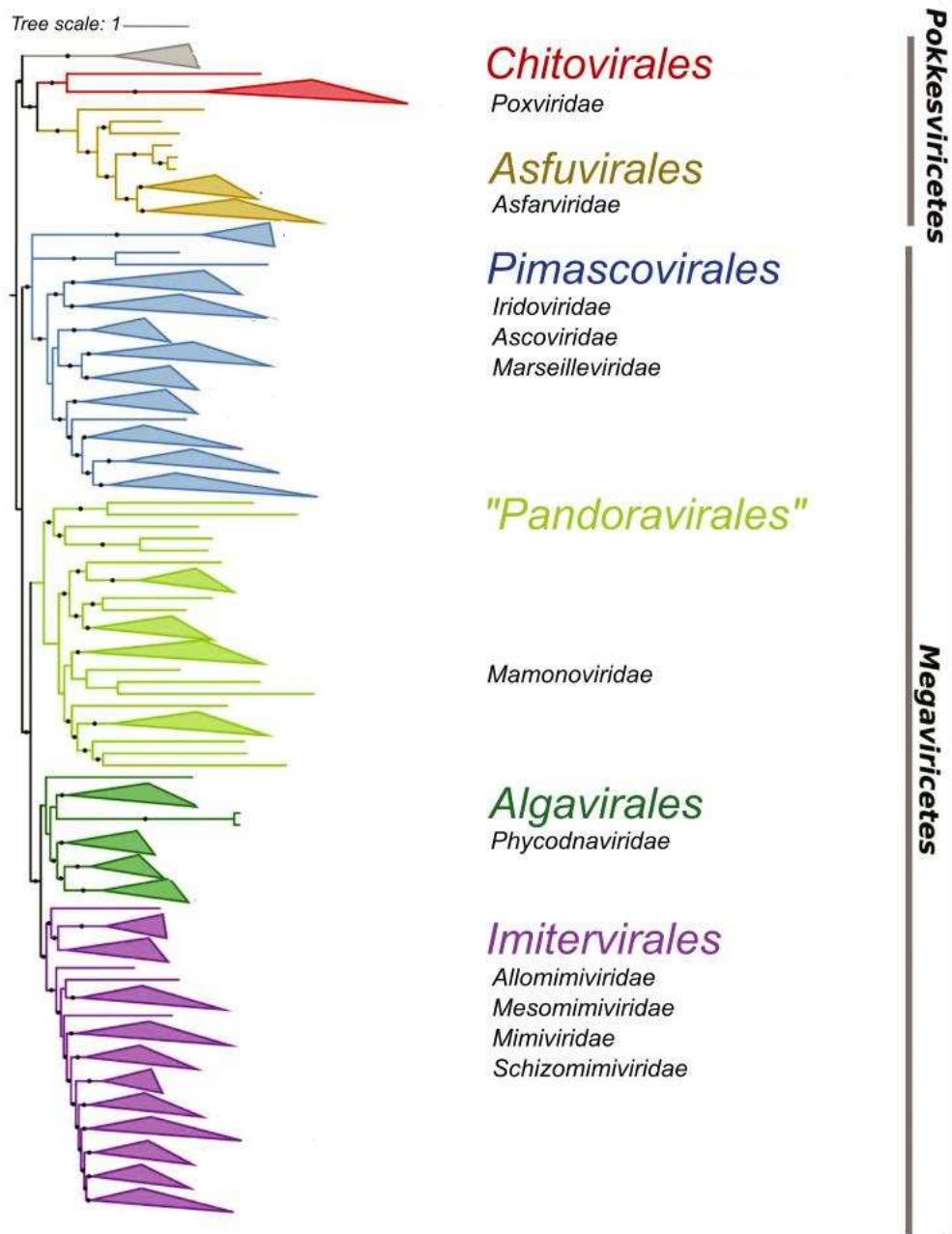
importante, em se tratando de vírus de interesse para a saúde humana e animal, para aprimorar a eficácia das intervenções médicas e de saúde pública através de métodos diagnósticos mais precisos, desenvolvimento de vacinas e descoberta de fármacos. No âmbito ambiental e puramente científico, a taxonomia viral desempenha um papel crucial ao proporcionar uma compreensão mais profunda e abrangente da diversidade viral, incluindo fenômenos como os vírus gigantes. Identificar e classificar esses vírus permite aos cientistas explorar os papéis ecológicos desses agentes virais, bem como compreender suas interações com outros organismos presentes no ambiente. A classificação taxonômica é essencial para investigar a evolução viral ao longo do tempo. Isso inclui a compreensão de eventos evolutivos únicos, como a descoberta de vírus gigantes, que desafiam as concepções tradicionais na virologia e acrescentam conhecimentos importantes para a biologia viral.

### 2.3 FILOGENIA E EVOLUÇÃO DE VÍRUS GIGANTES

A filogenia e evolução dos vírus gigantes têm sido objetos de intensa investigação. Atualmente, o filo *Nucleocytoviricota* (ou NCLDV) engloba onze famílias virais oficiais e diversos grupos e espécies taxonomicamente ainda não designados. Ainda hoje, a filogenia desse clado é bastante dinâmica e sujeita à novas alterações conforme novos isolados e sequências de vírus gigantes são descobertos. A construção da filogenia dos NCLDVs segue uma abordagem semelhante à rotineiramente utilizada para formas de vida celular, por meio da construção de árvores filogenéticas para genes marcadores universais, presentes em vírus desse grupo, como genes de capsídeo, DNA polimerase, ATPases e Helicases (Koonin & Yutin, 2019). No entanto, devido à escassez de genes conservados em todos os membros do grupo, a escolha de marcadores para a construção da filogenia é limitada. Ao analisar *clusters* de genes ortólogos (COG) presentes em todos os NCLDVs, Yutin et al. (2009) determinou um grupo *core* de 47 genes. Nos últimos anos, filogenias dos NCLDVs foram elaboradas para o crescente conjunto de genomas identificados, utilizando genes individuais quase-universais ou coleções de 3 a 7 genes (Koonin & Yutin, 2019). A árvore filogenética de NCLDVs mais recente (**figura 3**) consegue agrupar e definir clados monofiléticos com altos *branch support* indicando a criação de algumas famílias virais bem definidas nesse clado.



**Figura 3** – Árvore filogenética baseada em sete genes marcadores compartilhados entre todos os membros do filo *Nucleocyotoviricota* (NCLDV). Ordens virais representadas por diferentes cores.



Fonte: adaptado de Aylward et al., 2021.

O primeiro ramo na filogenia dos NCLDVs abrange as famílias *Allomimiviridae*, *Mesomimiviridae* e *Schizomimiviridae*, recentemente derivadas da família *Mimiviridae*. Há também a família *Phycodnaviridae* e *Mamonoviridae*, junto com os Pandoravírus, ainda sem família definida. Os *Mimiviridae* são conhecidos por seu tamanho excepcional, complexidade

genômica, e por terem sido os primeiros a serem identificados como gigantes em tamanho (Raoult et al., 2004). Incluem vírus como o APMV, que possui um genoma grande e uma variedade de genes relacionados à síntese de proteínas e ao metabolismo energético (Raoult et al., 2004). Os *Phycodnaviridae*, por sua vez, são predominantes em ambientes aquáticos e infectam algas, apresentando uma ampla gama de tamanhos genômicos (Wilson et al., 2009).

O segundo ramo compreende as famílias *Marseilleviridae*, *Iridoviridae* e *Ascoviridae*. Os *Marseilleviridae*, são caracterizados por sua relação com protistas e animais (Aherfi et al., 2014). Os *Iridoviridae* abrangem uma ampla gama de vírus, incluindo iridovírus, que infectam invertebrados, vertebrados (Chinchar et al., 2017). A família *Ascoviridae* é conhecida por conter vírus que parasitam insetos, e apresentam formas de vírion ovoides (Federici et al., 2009).

O terceiro ramo reúne as famílias *Poxviridae* e *Asfarviridae*, juntamente com os faustovírus, pacmanvírus, kaumoebavírus e AbALV. Os *Poxviridae* também são notáveis pela sua morfologia única, apresentando vírions em formato de tijolo (Condit et al., 2006). Incluem vírus como o vírus da varíola, que infecta mamíferos, e outros que afetam uma variedade de hospedeiros animais. Membros oficiais de *Asfarviridae*, por sua vez, são caracterizados por infecção em suínos e por apresentarem vírions com morfologia típica de icosaedro (Liu et al., 2019). Os faustovírus, pacmanvírus e kaumoebavírus foram identificados como filogeneticamente associados aos *Asfivirus*, levando à criação da terminação "*extended Asfarviridae*".

O reconhecimento das características notáveis dos vírus gigantes tem suscitado propostas conceituais inovadoras. Um dos conceitos mais proeminentes é a sugestão de que os vírus gigantes representam um "quarto domínio da vida", distinto, mas comparável aos três domínios celulares reconhecidos: bactérias, arqueias e eucariotos (Boyer et al., 2010). Esta proposta tem fundamentos na observação de características "semelhantes às células" dos vírus gigantes, referindo-se tanto ao tamanho dos vírions quanto ao genoma desses vírus. Entretanto, a contínua descoberta de novos isolados destaca que os vírus gigantes não são anomalias, mas sim entidades anteriormente negligenciadas devido a práticas de filtragem física estabelecidas. Essa observação sugere que, embora a ideia de um quarto domínio da vida seja uma abordagem conceitual útil para caracterizar o caráter "celular", em termos de tamanho, dos vírus gigantes, a classificação desses vírus como um domínio distinto pode ser errônea. Análises filogenômicas abrangentes indicam que a hipótese do quarto domínio não é sustentada (Yutin et al., 2014). Os vírus gigantes, apesar de seu tamanho incomum e complexidade genética, compartilham um conjunto central de genes que os vincula ao grupo NCLDV, esse por sua vez, geneticamente

associado a outros vírus na grande árvore filogenética viral (Bäckström et al., 2019; Yutin et al., 2014).

Existem duas propostas de modelos para a evolução dos vírus gigantes. Segundo o modelo redutivo, um genoma celular ancestral teria passado por um processo de redução em seu tamanho e conteúdo gênico, o que resultaria na dependência do genoma resultante em relação às células hospedeiras (Brandes & Linial, 2019). Uma evidência consistente com essa teoria é a presença de genes que desempenham funções celulares em praticamente todos os vírus gigantes. A outra proposta, mais aceita dentre os virologistas evolutivos, sustenta um modelo de expansão. De acordo com essa hipótese, os vírus gigantes atuais teriam se originado de vírus ancestrais menores, contendo apenas algumas dezenas de genes, e, por meio de duplicações gênicas e transferências horizontais de genes, teriam experimentado um rápido processo de expansão e diversificação (Filée & Chandler, 2010; Iyer et al., 2006; Krupovic & Koonin, 2015; Yutin et al., 2014). Pesquisas metagenômicas e a recente onda de descobertas de vírus gigantes, que indicam uma intensa troca gênica entre esses vírus e diversas formas de organismos compartilhando os mesmos ecossistemas, corroboram com esse modelo (Bäckström et al., 2019).

#### 2.4 EXTENDED *Asfarviridae*

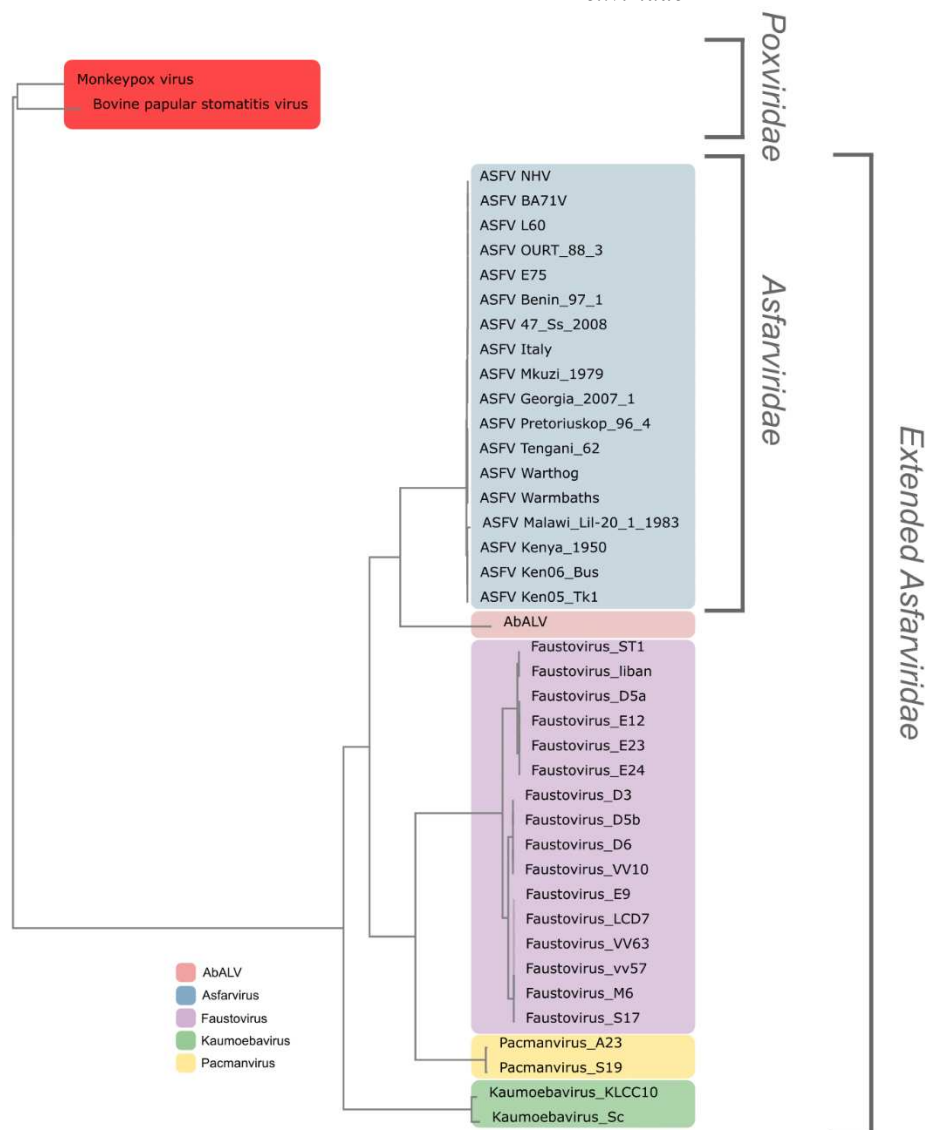
O clado "*extended Asfarviridae*" passou por uma trajetória evolutiva complexa desde a primeira inclusão do Vírus da Peste Suína Africana (ASFV) como parte da família *Iridoviridae* em 1984. Em 1998, genótipos associados ao ASFV receberam o gênero "*Asfivirus*" e foram designados para a nova família *Asfarviridae* (Alonso et al., 2018). Análises aprofundadas das estratégias de replicação e dos genes virais revelaram que o ASFV está relacionado a outros vírus gigantes de dsDNA originando o grupo dos NCLDVs (Iyer et al., 2001). Dentro de *Asfarviridae*, conforme novas espécies virais relacionadas filogeneticamente a essa família foram sendo identificadas, o grupo se expandiu de tamanho em número de indivíduos. Entretanto, até agora, somente espécies de *Asfivirus* compõem oficialmente essa família e, por isso, criou-se a denominação "*extended Asfarviridae*", ou "*Asfarviridae* estendida", em 2019 a fim de acomodar essas outras espécies temporariamente.

O crescimento significativo da família *Asfarviridae* ocorreu a partir de 2015, quando o início dos co-cultivos de isolados virais utilizando *Vermamoeba vermiformis* como hospedeiro, especialmente para faustovírus (Benamar et al., 2016). Esse método representou um afastamento do modelo clássico para vírus gigantes de *Acanthamoeba* spp., marcando um passo significativo na compreensão da diversidade dessa família viral.

Portanto, além do ASFV, agora o grupo *extended Asfarviridae* abrange uma variedade de vírus, como faustovírus, kaumoebavírus, pacmanvírus e o AbALV, que, de certa forma, compartilham uma proximidade genética dos ASFV (**figura 4**). Especula-se que esses quatro grupos virais possuem um ancestral em comum e que os kaumoebavírus foram os primeiros a divergir (Andreani et al., 2017).

Anteriormente limitada ao ASFV, a família *Asfarviridae* expandiu-se para incluir outras espécies virais que infectam diferentes hospedeiros e apresentam variações genômicas notáveis. A importância desse clado, tanto em termos de saúde animal quanto de entendimento da diversidade viral gênica, destaca a necessidade contínua de pesquisas e estudos com o clado "*extended Asfarviridae*".

**Figura 4** – Árvore filogenética de *extended Afarviridae* baseada no gene da DNA polimerase B. A raiz da árvore está baseada em vírus de *Poxviridae*.



Fonte: elaborado pelo autor (2024)

### 2.4.1 *Asfivirus*

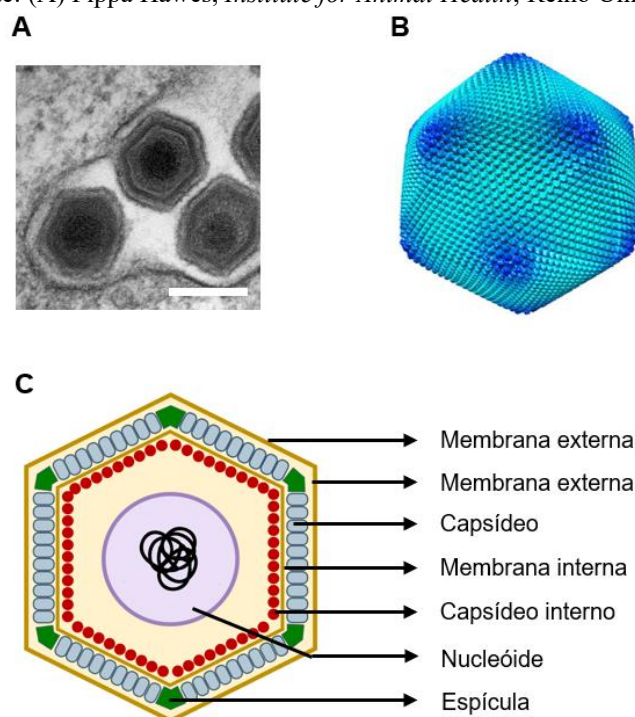
O ASFV, pertencente ao gênero *Asfivirus* e membro fundador da família *Asfarviridae*, é o causador da Peste Suína Africana (ASF), uma doença devastadora para suínos. Sua história remonta à década de 1920, quando foi inicialmente descrita no Quênia (Eustace Montgomery, 1921). Atualmente, a ASF é considerada uma das mais sérias ameaças para a indústria suína global, especialmente para pequenos produtores, devido à sua alta contagiosidade e alta taxa de mortalidade, atingindo tanto suínos domésticos quanto selvagens (Berthe, 2020).

A análise da estrutura de seu vírion por microscopia eletrônica (**figura 5A**) revelou uma partícula com aproximadamente 200 nm de diâmetro enquanto que a microscopia crioelétrica revela com detalhes a configuração icosaédrica de seu capsídeo (**figura 5B**) (Andrés et al.,

2020; Liu et al., 2019). O vírion do faustovírus é formado por uma morfologia em multicamadas composta por vários domínios: o núcleo interno formado pelo genoma central contém o nucleóide, que é revestido por uma espessa camada proteica denominada matriz ou involucro interno do núcleo; uma membrana lipídica interna envolvendo o núcleo e, por fim, o capsídeo icosaédrico, que constitui a camada mais externa dos vírions intracelulares (**figura 5C**). Essa configuração icosaédrica é compartilhada pelos outros membros do clado, como o faustovírus, kaumoebavírus, pacmanvírus e AbALV (Klose et al., 2016; Liu et al., 2019; Matsuyama et al., 2023). Os vírions extracelulares possuem uma camada lipídica externa adicional obtida quando o vírus brota através da membrana plasmática da célula hospedeira (**figura 5C**) (Andrés et al., 1997; G. Li et al., 2020; Liu et al., 2019).

**Figura 5** – Estrutura do vírion do ASFV. (A) Micrografia eletrônica do vírion, barra representa 200 nm (B) Reconstrução cryo-EM do capsídeo do ASFV cepa BA71V, colorida de acordo com a distância radial do centro do vírus (C) Representação esquemática e simplificada da estrutura multicamada do vírion do ASFV.

Fonte: (A) Pippa Hawes, *Institute for Animal Health*, Reino Unido, Disponível em:



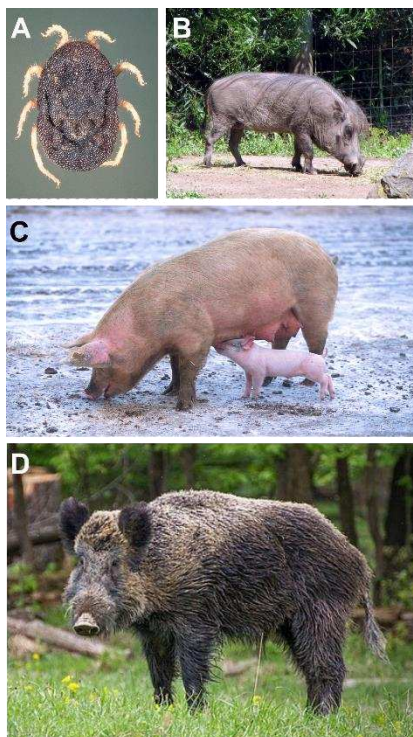
<https://ictv.global/report/chapter/asfarviridae/asfarviridae>. (B) Andrés et al., 2020, (C) elaborado pelo autor (2024).

O genoma do ASFV é composto por uma molécula linear de dsDNA, com um tamanho variando entre 170 e 190 kb, que pode resultar na codificação de mais de 150 *Open Reading Frames* (ORFs) (Kabra et al., 2020). A variação no tamanho do genoma entre diferentes isolados resulta do ganho ou perda de cópias de genes pertencentes às famílias multigênicas e à variação no número de repetições em tandem em regiões não codificantes do genoma do ASFV (Chapman et al., 2008; Dixon et al., 2013). Sua complexidade genômica é evidenciada

pela presença de sequências terminais invertidas e complementares, bem como por blocos de unidades repetitivas de 2,1 kb adjacentes às extremidades. Essas ORFs estão densamente compactadas, presentes em ambas fitas de DNA, com distâncias intergênicas geralmente inferiores a 200 nt, algumas contendo pequenas sequências repetidas em tandem (Alonso et al., 2018).

A transmissão do ASFV ocorre principalmente por intermédio do carrapato (*Ornithodoros* spp.) (**figura 6A**), além do contato direto de animais saudáveis com animais infectados. É importante destacar que esse vírus é o único arbovírus de DNA conhecido (Cwynar et al., 2019). Três ciclos de transmissão foram identificados em áreas endêmicas: (i) ciclo entre o javali-africano (*Phacochoerus africanus*) (**figura 6B**) e o carrapato, mais comum na África; (ii) ciclo do suíno doméstico (*Sus domesticus*) (**figura 6C**), onde o ASFV é transmitido para esses porcos através de carrapatos na África e Europa e (iii) o ciclo do porco doméstico (*Sus domesticus*) com o Javali (*Sus scrofa*) (**figura 6D**), onde o ASFV é diretamente transmitido entre esses animais, sem a necessidade da presença do carrapato (Kabra et al., 2020). Além disso, outras rotas de transmissão incluem a ingestão de subprodutos animais contaminados, contato com suínos infectados, tecidos, carcaças ou picadas de carrapatos infectados, importação de produtos suínos infectados e contaminação por fômites (Gaudreault et al., 2020).

**Figura 6** – Hospedeiros conhecidos do ASFV. (A) Carrapato (*Ornithodoros* spp.), vetor. (B) Javali Africano (*Phacochoerus africanus*), (C) Porco doméstico (*Sus domesticus*), (D) Javali (*Sus scrofa*).



Fonte: Wikipédia.

O ASFV demonstra afinidade por células do sistema mononuclear-fagocitário, como macrófagos e linhagens específicas de células reticulares. A entrada do vírus nas células é predominantemente mediada por endocitose mediada por clatrina, macropinocitose e fagocitose (Basta et al., 2010; Galindo et al., 2015; Sánchez et al., 2012). O processo de desencapsulação do ASFV depende de fatores do hospedeiro encontrados durante sua fusão endossomal, com a desmontagem do capsídeo ocorrendo no pH ácido do interior endossomal (Galindo & Alonso, 2017). Após a degradação do capsídeo e a fusão da membrana, os núcleos virais são liberados dos endossomos. A síntese inicial de mRNA tem início imediatamente após a entrada, utilizando enzimas e fatores empacotados no núcleo viral (Galindo & Alonso, 2017). A replicação e montagem do DNA viral ocorrem em áreas perinucleares, nas "fábricas virais", bem comum em NCLDVs, próximas ao centro organizador de microtúbulos após um breve período de replicação no núcleo (Hernaiz et al., 2006; Quetglas et al., 2012; Simões et al., 2019). A saída do vírus ocorre por transporte microtubular até a superfície celular, onde a liberação viral por brotamento resulta em vírions extracelulares cobertos por um envoltório lipídico adicional (Jouvenet et al., 2004; Salas & Andrés, 2013).

A infecção por ASFV em suínos resulta em alta morbidade e mortalidade, com taxas de mortalidade que podem chegar a 100% (Penrith & Vosloo, 2009). Os sinais clínicos variam de formas hiperagudas a crônicas. A capacidade do vírus de eficientemente induzir citopatologia em macrófagos é um fator crítico para a virulência. Suínos que sobrevivem à infecção podem permanecer persistentemente infectados, sem sinais clínicos, por um período ainda desconhecido (Kabra et al., 2020).

Atualmente, a ASF é endêmica em grande parte da África, afetando cerca de 35 países, incluindo Madagascar (Penrith et al., 2019). Até 1999, a presença deste vírus estava restrita à África; entretanto, em 2007, a partir da Geórgia, o vírus se disseminou para nações da Europa Oriental, Cáucaso e Federação Russa, ocasionando surtos expressivos. Na União Europeia, Lituânia e Polônia foram os primeiros a notificar casos em javalis em 2014, seguidos pela propagação para Estônia, Letônia e Lituânia (Gaudreault et al., 2020). Em 2018, o ASFV também afetou a China, o maior produtor de carne suína mundial. Após isso, mais 8 países asiáticos foram afetados, resultando na morte de mais de 5 milhões de suínos (Gaudreault et al., 2020). O Brasil, e outros países da América, como o México, também já foram afetados pelo ASFV. Aqui, um surto aconteceu entre 1978 e 1979 e, após um intenso programa de erradicação organizado pelo governo brasileiro, incluindo o extermínio de mais de 66 mil suínos, o país foi declarado zona livre de ASF em 1984 (De Paula Lyra et al., 1986). Mais de 60 países em todo



o mundo já apresentaram surtos de ASF desde a primeira identificação do vírus em 1921, e os surtos globais desde 2018 apresentaram aumento de 25 % (*A global battle against African swine fever - Global Times*, 2019).

O controle da ASF é desafiador devido à complexidade dos ciclos de manutenção do vírus, à persistência ambiental e à falta de vacinas eficazes. A erradicação é complicada pela alta taxa de transmissão e pelas possíveis reintroduções do vírus em áreas livres de ASF. ASF representa uma séria ameaça para a segurança alimentar global, com implicações significativas na economia e na saúde animal. Estratégias integradas de vigilância, biossegurança e colaboração internacional são essenciais para conter e erradicar a propagação desse vírus altamente contagioso.

#### 2.4.2 Faustovírus

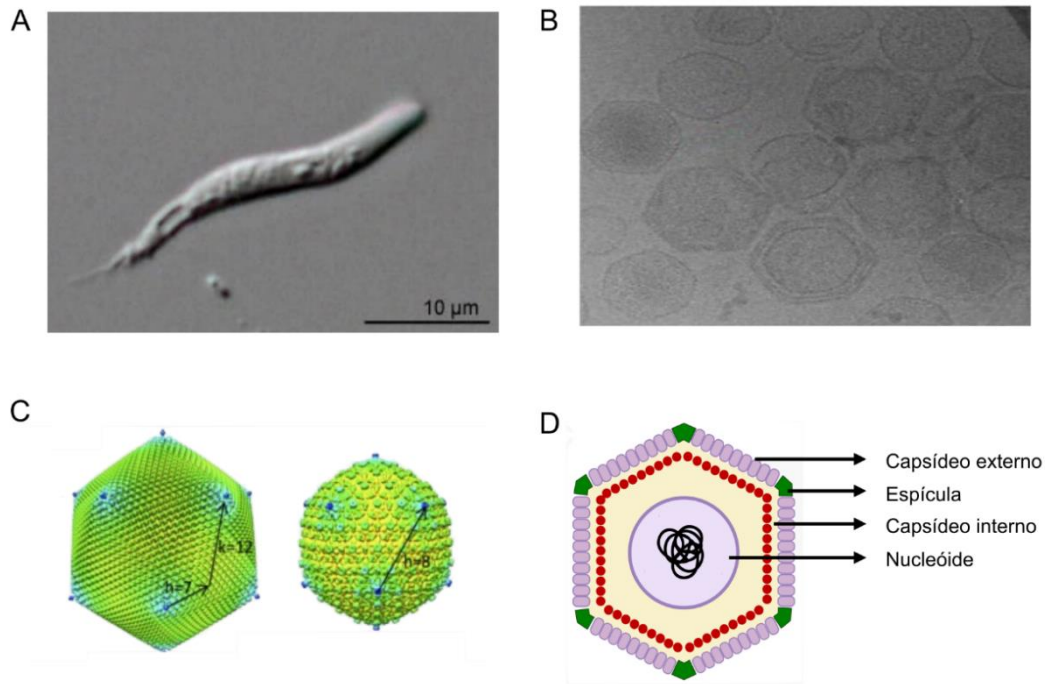
Descoberto em 2015 a partir de amostras de águas residuais na França e Senegal (Reteno et al., 2015), o faustovírus foi isolado a partir de estudos conduzidos em *V. vermiformis* (**figura 7A**). Além disso, esse vírus já foi detectado em mosquitos hematófagos *Culicoides imicola* e em seus hospedeiros roedores, assim como em amostras humanas, sugerindo uma ampla distribuição (Temmam et al., 2015). Como *V. vermiformis* é amplamente encontrada no ambiente humano, incluindo redes de água hospitalar (Pagnier et al., 2015; Thomas et al., 2006), água potável (A. Marouf Et Al., 2021; Javanmard et al., 2017), fezes humanas (Bradbury, 2014), *swabs* nasais (Arab-Mazar et al., 2019) e lentes de contato de pacientes com ceratite (ABEDKHOJASTEH et al., 2013; HAJIALILO et al., 2015; Kennedy et al., 1995), levantam-se discussões sobre a possível associação desse vírus com os seres humanos (Rolland et al., 2019). Apesar dessa associação, desconhece-se efeitos patogênicos do faustovírus em humanos; mais pesquisas são necessárias para determinar o modo de infecção e as possíveis consequências dessa infecção, se existirem.

O genoma do faustovírus é composto por dsDNA, geralmente ultrapassando 450 kb, abrigando mais de 450 ORFs, com um conteúdo de G+C compreendido entre 36,2% e 39,6% (Reteno et al., 2015). O faustovírus possui um genoma com formato circular, exceto pelo isolado *Faustovirus liban*, para o qual foi identificado um genoma linear (Benamar et al., 2016). Apesar de ser classificado como um vírus pertencente ao clado NCLDV, os faustovírus compartilham menos de um quarto de seus genes com outros NCLDVs; no entanto, aproximadamente 46% são homólogos a genes bacterianos, enquanto o restante consiste em genes órfãos (ORFans) – sem homologia em bancos de dados (Benamar et al., 2016). Destaca-se também a presença significativa de genes parálogos (com múltiplas cópias no mesmo

genoma). Outra característica que o diferencia é uma região de 17 kb no genoma do Faustovírus que abriga cinco ORFs, responsáveis pela codificação de polipeptídeos relacionados à *Major Capsid Protein* (MCP), incorporando íntrons e éxons, diferenciando-se notavelmente da estrutura observada em outros vírus em termos da presença desses elementos não codificantes, os quais são raros em espécies virais (Klose et al., 2016). Por exemplo, em *Asfivirus*, o gene que codifica a MCP é um único fragmento genômico de cerca de 2 kb. No faustovírus, entretanto, a MCP é codificada por 13 éxons separados por 12 íntrons extensos (Klose et al., 2016). Posteriormente, também foi identificada a presença de íntrons tipo-I na sequência genômica de kaumoebavírus (Bajrai et al., 2016).

O vírion do faustovírus exibe um tamanho de 200–240 nm (**figura 7B**), com configuração icosaédrica apresentando (**figura 7C**), pela primeira vez no grupo, um capsídeo proteico de dupla camada (**figura 7C**) (Klose et al., 2016). Sua estrutura lembra a dos *Asfivirus*, mas sem a presença da membrana externa e a substituição da membrana interna pelo capsídeo interno (**figura 7D**). A configuração de dupla camada proteica a fim de encapsidar o genoma surge como uma novidade para vírus de DNA, usualmente os vírus possuem uma combinação de camada proteica com ou sem membrana lipídica. Em vez disso, a membrana interna que geralmente encapsula o genoma viral em muitos outros NCLDVs foi substituída por uma camada adicional de proteína. É necessário investigar, com base em novas identificações virais, se esse fenômeno é exclusivo do faustovírus ou se representa um exemplo de evolução divergente. Neste último cenário, alguns membros dos NCLDVs poderiam ter adquirido uma membrana para salvaguardar seu material genético, enquanto outros desenvolveram uma camada de proteção secundária fundamentada em proteínas (Klose et al., 2016).

**Figura 7** – Faustovírus e seu hospedeiro conhecido. (A) *Vermamoeba vermiformis*, protozoário hospedeiro dos faustovírus. (B) Micrografia eletrônica do vírion do Faustovirus E12, (C) Reconstrução cryo-EM do capsídeo, colorida de acordo com a distância radial do centro do vírus, (D) Representação esquemática e simplificada da estrutura multicamada do vírion.



Fonte: (A)Wikipédia, (B,C) Klose et al., 2016, (D) elaborado pelo autor (2024).

De acordo com Reteno et al. (2015), o ciclo replicativo do faustovírus, com duração de 18 a 20 horas, inicia-se com a ameba ingerindo partículas virais individuais por meio de fagocitose. Cerca de 2 a 4 horas após a infecção, as partículas virais são internalizadas por meio de vacúolos fagocíticos. Embora as partículas apareçam próximas ao núcleo do hospedeiro, não há evidências de que o vírus esteja dentro do núcleo ou interaja com a membrana nuclear. Similarmente ao mimivírus, no qual um canal é criado para a passagem de proteínas e DNA das partículas, as partículas do faustovírus esvaziam seus compartimentos internos no citoplasma da ameba com a fusão da membrana lipídica interna que delimita o núcleo da partícula com a membrana do vacúolo. A célula hospedeira passa por uma reorganização, de modo que, de 8 a 10 horas após a infecção, há o surgimento da fábrica viral no citoplasma. Entre 12 e 18 horas após a infecção, a fábrica viral ocupa todo o citoplasma, que está completamente preenchido com novas partículas virais. Entre 18 e 20 horas após a infecção, as partículas virais são liberadas por lise celular.

### 2.4.3 Kaumobavírus

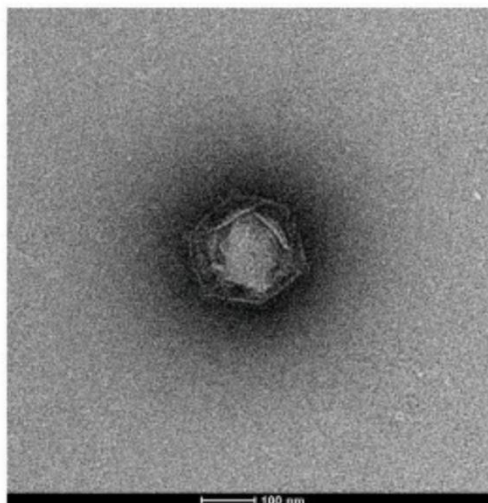
O segundo grupo de vírus gigante isolado após o início de estudos de co-cultura em *V. vermiformis* (figura 7A) foi o kaumobavírus, também uma nova adição ao clado de *extended*

*Asfarviridae*. Este vírus gigante foi inicialmente identificado em amostras de águas residuais na região sul de Jeddah, na Arábia Saudita, em 2016, por pesquisadores da *King Abdulaziz University*, de onde origina-se o nome desse vírus (Bajrai et al., 2016). Posteriormente, uma segunda cepa denominada LCC10 também foi isolada a partir de águas residuais no sul da França, ampliando assim nosso conhecimento sobre a diversidade biológica e distribuição desse grupo de vírus.

O genoma do kaumoebavírus, com uma extensão entre 350 e 362,6 kb, revelou-se como uma molécula de dsDNA linear, contendo mais de 450 ORFs, com conteúdo G+C global de 43,7%, apresentando terminações em formato de *hairpin* covalentemente fechadas, diferenciando-se dos faustovírus (Bajrai et al., 2016). Observou-se que entre 67% e 76% do conteúdo gênico do kaumoebavírus são ORFans, sem correspondência alguma com outras proteínas no banco de dados do *non-redundant database* (nr) do NCBI (Hannat et al., 2023). Há ainda, assim como nos faustovírus, a presença de regiões intrônicas no genoma desse vírus, incluindo a região correspondente ao gene *MCP*. Em regiões intergênicas de seu genoma, percebe-se uma preferência por sequências ricas em AT, especialmente os *motifs* TATTT e TATATA, compartilhando essa característica com os faustovírus (Oliveira et al., 2018). Esses *motifs* desempenham um papel crucial como promotores, reforçando a ideia de uma possível origem ancestral comum entre esse grupo de vírus gigantes. Por fim, há um viés de codificação gênica em de uma das fitas de seu DNA para ao menos dois terços de seu genoma, característica ainda não observada em outros membros de *extended Asfarviridae*. Essa observação foi feita nos dois isolados de kaumoebavírus, indicando um surgimento ancestral em sua linhagem ou até mesmo em *Asfarviridae* e posteriormente perdido (Geballa-Koukoulas, Andreani, et al., 2021).

A estrutura física do kaumoebavírus ainda não foi intensamente estudada. O que sabe-se, através de microscopia eletrônica, é que seus vírions são partículas icosaedrais possuindo um conformação em multi-camada, apresentando diâmetro de aproximadamente 250 a 260 nm (**figura 8**) (Bajrai et al., 2016; Geballa-Koukoulas, Andreani, et al., 2021).

**Figura 8** – Micrografia eletrônica do vírion do Kaumoebavirus KLCC10.



Fonte: Geballa-Koukoulas.

No âmbito da replicação, o ciclo replicativo do kaumoebavírus, conta com a mesma duração que a dos faustovírus, entre 16 e 20 horas, e envolve as mesmas etapas, desde a fagocitose, fusão de membranas, fábrica viral e a liberação das partículas por meio da lise celular (Bajrai et al., 2016; Geballa-Koukoulas, Andreani, et al., 2021). Notoriamente, ao contrário do faustovírus, o kaumoebavírus não induz alterações morfológicas no núcleo da célula hospedeira durante seu ciclo replicativo (Bajrai et al., 2016; Geballa-Koukoulas, Andreani, et al., 2021).

#### **2.4.4 Pacmanvírus**

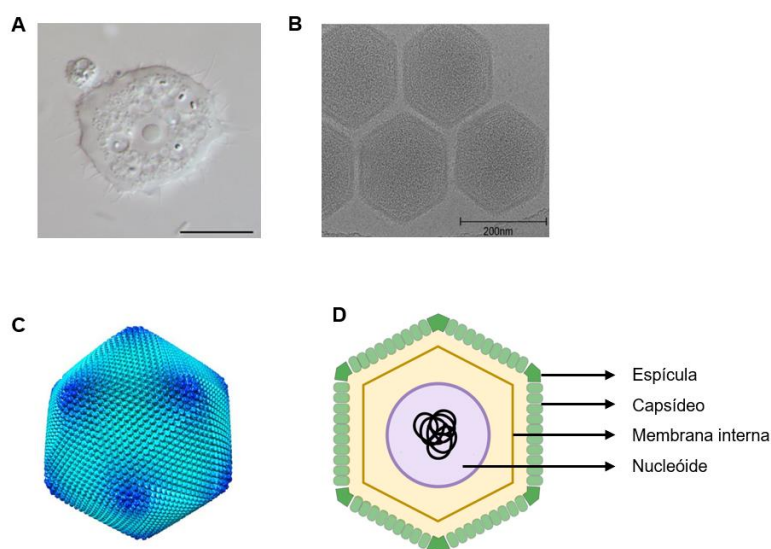
A descoberta do pacmanvírus, recentemente isolado de co-cultura com *A. castellanii*, amplia significativamente nosso entendimento sobre o clado *extended Asfarviridae*. Este novo vírus, denominado Pacmanvirus A23, foi isolado pela primeira vez em 2017 a partir de amostra ambiental coletada na Argélia (Andreani et al., 2017). O segundo exemplar desse grupo, o Pacmanvirus S19, foi isolado de águas residuais, também na Argélia (Geballa-Koukoulas, Abdi, et al., 2021).

O genoma dsDNA linear do Pacmanvirus A23 abrange 395 kb, menor somente que os faustovírus, com um conteúdo G+C de 33,62%. Distingue-se por uma alta densidade de codificação, compreendendo por volta de 465 ORFs e ausência de grandes regiões repetitivas (Andreani et al., 2017). O pacmanvírus é o único vírus desse clado que contém gene que codifica tRNA (ile), provavelmente para se adaptar a alguma necessidade traducional durante a infecção (Dreher, 2010). Com relação ao seu conteúdo gênico codificante, o pacmanvírus exibe uma proporção de 52.5% de genes ORFans, sem correspondência com nenhum registro

no banco de dados do nr (Andreani et al., 2017). A presença de parálogos no genoma do pacmanvírus foi determinada, indicando poucas famílias de genes parálogos, caso semelhante ao do faustovírus, sugerindo uma limitação espacial ainda presente no genoma (Andreani et al., 2017).

A estrutura física de seu vírion é caracterizada por partículas icosaédricas irregulares, apresentando uma média de 175 nm (**figura 9A**) (Andreani et al., 2017). A análise tridimensional por microscopia crioeletrônica revela uma arquitetura icosaedral (**figura 9B**) semelhante à dos outros membros de seu clado, porém com a distinção de apresentar uma membrana lipídica interna em vez de uma proteica (**figura 9C**), como nos faustovírus (Andreani et al., 2017). O *locus* genômico responsável pelo MCP assemelha-se ao dos *Asfivirus*, sendo codificado por uma única ORF, sem a presença de íntrons (Andreani et al., 2017). Algumas das primeiras micrografias eletrônicas revelaram estruturas capsidiais defeituosas o que inspiraram o nome pacmanvírus, se referindo ao personagem de *video game*.

**Figura 9** – Estrutura do Pacmanvirus e seu hospedeiro conhecido. (A) *Acanthamoeba castellanii*, hospedeiro do Pacmanvirus, (B) Micrografia eletrônica do vírion do Pacmanvirus A23, (C) Reconstrução cryo-EM do capsídeo, colorida de acordo com a distância radial do centro do vírus, (D) Representação esquemática e simplificada da estrutura multicamada do vírion.



Fonte: (A) Wikipédia; (B-C) Andreani et al., 2017; (D) elaborado pelo autor (2024).

O Pacmanvirus demonstra uma replicação excepcionalmente rápida em *A. castellanii*, com a lise da ameba completando em apenas 8 horas após a entrada do vírion. O ciclo infeccioso envolve as mesmas etapas dos faustovírus e kaumobavírus, incluindo a fagocitose do vírion, fusão de membranas, surgimento da fábrica viral no citoplasma e a liberação das partículas por

meio da lise celular da ameba, a única possível diferença é uma provável interação com a mitocôndria, ainda a ser provada (Andreani et al., 2017).

#### 2.4.5 *Abalone asfar-like virus* (AbALV)

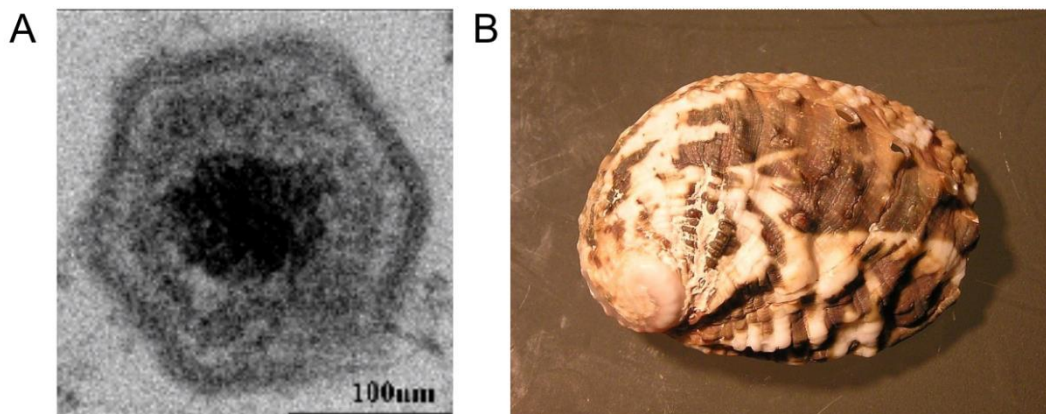
Abalones são moluscos gastrópodes marinhos pertencentes ao gênero *Haliotis*, conhecidos por suas conchas espiraladas com interior iridescente, encontrados em diversas regiões costeiras ao redor do mundo e utilizada na alimentação humana. O *Abalone asfar-like virus* (AbALV), designado provisoriamente dessa maneira, foi identificado como o agente etiológico responsável por mortalidade em massa entre esses moluscos (Matsuyama et al., 2020). A doença, denominada amiotrofia de abalone, foi inicialmente reconhecida na década de 1980 (Nakatsugawa et al., 1988), porém os esforços para identificar um agente causador não obtiveram resultados na época. A amiotrofia de abalone resulta em mortalidade em massa de abalones jovens, apresentando sintomas iniciais como atrofia muscular nos pés e alterações na pigmentação da concha. Esta doença tem um impacto significativo na indústria pesqueira, especialmente na Ásia, e afeta também a produção de madrepérola, obtida das conchas de abalone (Matsuyama et al., 2021).

A recente descoberta desse vírus ocorreu por meio de sequenciamento de nova geração (NGS) e pesquisas epidemiológicas baseadas em PCR. Amostras de abalones doentes foram coletadas em instalações de cultivo onde ocorreram mortalidades em massa (Matsuyama et al., 2020).

O genoma do AbALV possui um tamanho de 281 kb e mais de 300 ORFs preditas. É um genoma de dsDNA linear, com conteúdo G+C de 31,32%, o menor de *extended Asfarviridae*, e não codifica tRNAs, ao contrário dos Pacmanvirus. Análises de sua estrutura genômica e filogenias o colocaram como pertencente ao clado *extended Asfarviridae* (Matsuyama et al., 2023).

A morfologia das partículas virais do AbALV ainda não é totalmente resolvida, porém, imagens de microscopia eletrônica revelam uma estrutura semelhante à do ASFV (Matsuyama et al., 2023). Apresentam uma forma icosaédrica com aproximadamente 200 nm de tamanho (**figura 10**), composta por nucleóide, um provável capsídeo interno e o capsídeo externo. Assim como todos os NCLDVs, partículas virais são observadas no citoplasma das células infectadas. Mas, apesar de ser comum aos outros membros NCLDVs, fábricas virais não foram observadas no citoplasma.

**Figura 10** – AbALV e seu hospedeiro conhecido. (A) Micrografia eletrônica do vírion do AbALV. (B) *Haliotis madaka*, uma das espécies de abalone que servem como hospedeiro para AbALV.



Fonte: (A) Matsuyama et al., 2023, (B) Wikipedia.

## 2.5 CONCEITOS FUNDAMENTAIS DE PANGENOMA

A complexidade da diversidade biológica de uma espécie demanda uma quantidade significativa de informações genômicas. O aumento da quantidade de DNA sequenciado e a evolução das tecnologias de sequenciamento, com consequente redução do tempo e o custo de sequenciamento, tem resultado em um aumento exponencial no número de genomas disponíveis em bancos de dados. Isso tem dado origem a novas áreas de pesquisa, como a genômica comparativa, que explora campos como a evolução genômica. Nesse contexto, o termo "pangenoma" foi cunhado por Tettelin et al. (2005), especificamente para descrever o conteúdo gênico total de várias cepas de *Streptococcus agalactiae*, principal causador de infecção neonatal em humanos. Este conceito revela-se fundamental na genômica comparativa, que busca comparar o conteúdo genético de micro-organismos taxonomicamente relacionados.

Tettelin sugere que o pangenoma apresenta uma estrutura estratificada, dividida em três subgrupos, que desempenham um papel fundamental na compreensão da complexidade e da diversidade genômica: genoma central ou *core*, genoma acessório e genes *singleton*, ou específicos da cepa (**figura 11**). O grupo dos genes *core* abriga o núcleo dos genes essenciais, compartilhados por todas as cepas analisadas, muitos dos quais desempenham papéis vitais na sobrevivência do organismo. Esse conjunto é majoritariamente formado pelos genes chamados de *housekeeping*, genes que mantêm a maquinaria essencial de funções metabólicas e genômicas.

O genoma acessório, muitas vezes de forma controversa chamado de “genoma dispensável”, é composto por genes presentes em mais genomas, mas não em todos. Esse conjunto de genes é altamente variável e frequentemente associado a adaptações ambientais

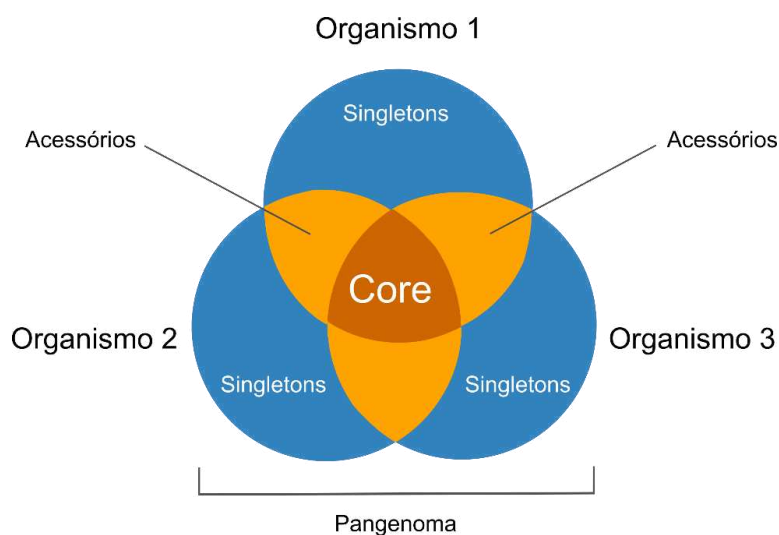


específicas, representando características adquiridas por cada espécie ao longo de sua evolução. Esses genes desempenham um papel crucial na capacidade de uma espécie responder e se adaptar a diferentes ambientes, constituindo uma rica fonte de variabilidade genética que confere flexibilidade à população de um organismo para explorar ambientes diversos.

A distribuição dos genes no genoma acessório pode ser classificada em diferentes categorias, como *soft core*, *shell* e *cloud*, de acordo com sua presença nas espécies. O *soft core* engloba os genes presentes em pelo menos 90% das espécies, indicando uma certa estabilidade e importância funcional compartilhada por uma grande parte das espécies do grupo, mas não tão necessária a ponto de estar em todos os genomas. Por sua vez, o grupo *shell* compreende os genes presentes em menos de 90% das espécies, porém em mais de duas espécies, refletindo uma diversidade intermediária e uma adaptabilidade a ambientes variáveis. Já o *cloud* consiste nos genes presentes em uma ou duas espécies apenas, sugerindo uma variação genética mais restrita e específica.

Dentro do grupo dos genes *cloud* há a presença dos genes *singleton*, os quais representam uma categoria especial, sendo genes presentes exclusivamente em uma única cepa e muitos dos quais também podem ser genes ORFans, ou seja, sem qualquer homologia ainda descoberta nos bancos de dados (Siew & Fischer, 2003). Esses genes singulares também podem desempenhar papéis em características específicas da cepa, conferir vantagens seletivas em determinados contextos, assim como os acessórios. São muitas vezes resultado de eventos evolutivos únicos como transferência gênica horizontal (HGT), evolução devido ao acúmulo de mutações em genes pré-existentes ou até mesmo ao processo de surgimento de genes *de novo* (Costa et al., 2020). O surgimento de genes *singletons* pode estar relacionado à adaptação de um organismo a pressões ambientais novas ou variáveis (Ezzatpour et al., 2023).

**Figura 11** - Diagrama de Venn ilustrando os subgrupos de um pangenoma formado por três organismos.



Fonte: elaborado pelo autor (2024)

Ao buscar genes homólogos, a análise pangenômica desempenha um papel fundamental na compreensão da evolução genômica, categorizando-os em dois subgrupos distintos: ortólogos e parálogos. Os genes ortólogos são caracterizados por compartilharem uma origem ancestral comum que existia antes da especiação. Eles mantêm funções biológicas equivalentes em diferentes organismos ou cepas, refletindo uma conservação significativa ao longo do tempo evolutivo. Essa conservação de função biológica sugere que os ortólogos desempenham papéis essenciais para a sobrevivência e adaptação dos organismos em seus respectivos ambientes (Koonin, 2005; Padró et al., 2022; Willger et al., 2012). Em contraste, os genes parálogos têm sua origem em duplicações gênicas, eventos que ocorrem durante a evolução e resultam na presença de múltiplas cópias de um gene na mesma espécie ou linhagem (Fitch, 1970; Koonin, 2005). Essas cópias, ou parálogos, são propensas a sofrer mutações ao longo do tempo, gerando variação em suas sequências e, conseqüentemente, em suas funções biológicas (Fitch, 1970; Koonin, 2005).

A análise pangenômica, com um foco particular nos produtos gênicos, envolve o emprego de ferramentas de bioinformática avançadas que realizam a clusterização das proteínas de acordo com sua similaridade. Os algoritmos, como o BLAST (*Basic Local Alignment Search Tool*) e similares, destacam-se como os mais comuns (Vernikos, 2020). Eles agrupam proteínas ao considerar similaridades em suas sequências de aminoácidos por meio de alinhamentos todos-vs-todos par-a-par (Camacho et al., 2009). Outro método importante é o uso de grafos, como o algoritmo *Markov Cluster*, que agrupa proteínas com base em suas interações representadas como grafos. Nessa abordagem, os nós representam entidades, como proteínas, e

as arestas representam relações ou interações entre essas entidades (W. Li et al., 2012). Há ainda o uso de algoritmos que realizam a contagem de palavras em sequências, sejam elas de nucleotídeos ou aminoácidos. Esses algoritmos, como exemplificado pelo CD-HIT, utilizam a contagem de palavras para avaliar similaridades entre as sequências, sem a necessidade de realizar alinhamentos diretos (Fu et al., 2012). O agrupamento de proteínas facilita a identificação de famílias de genes, proporcionando uma compreensão das relações evolutivas. Além disso, ela fornece informações cruciais sobre homologias, identificação de ortólogos e parálogos.

Conforme sugerido por Tettelin et al. (2008), um pangenoma pode ser classificado como aberto ou fechado: em um pangenoma aberto, seu tamanho (em genes) aumenta indefinidamente ao adicionar novos genomas; assim, a adição de novos genomas provavelmente resultará em genes inéditos. Por outro lado, em um pangenoma fechado, a inclusão de novos genomas não resultará na descoberta de novos genes. Essa categorização pode ser realizada através da aplicação da Lei de Heap e de uma lei de potência. A Lei de Heap é expressa como:

$$n = \kappa N^\gamma$$

Onde  $n$  representa o tamanho do pangenoma,  $N$  é o número de genomas utilizados, e  $\kappa$  e  $\gamma$  são os parâmetros de ajuste. O expoente  $\gamma$  é crucial para determinar se o pangenoma apresenta uma característica aberta. Quando  $\gamma < 0$ , o pangenoma é considerado fechado, e seu tamanho se aproxima de uma constante à medida que mais genomas são incorporados. Por outro lado, quando  $\gamma > 0$ , o pangenoma é classificado como aberto, e seu tamanho aumenta com a inclusão de mais genomas. A propriedade aberta de um pangenoma específico também pode ser estimada utilizando um modelo de lei de potência com base na fórmula:

$$\Delta n = \kappa N^{-\alpha}$$

Em que  $\Delta n$  é o número de genes recém-adicionados,  $N$  é o número de genomas utilizados, e  $\kappa$  e  $\alpha$  são os parâmetros de ajuste. Se  $\alpha > 1$ , o pangenoma é considerado fechado; enquanto  $\alpha < 1$ , o pangenoma é considerado aberto. A Lei de Heap e a lei de potência compartilham similaridades matemáticas.

Exemplificando a importância da análise pangenômica, os primeiros pangenomas a serem implementados, centrados em bactérias, identificaram genes variáveis relacionados a virulência bacteriana e a vias biológicas não essenciais (Tettelin et al., 2005). A partir disso, houveram estudos subsequentes que, ao aplicar a abordagem pangenômica, realizaram a descoberta de vias bioquímicas novas, como sistemas de defesa (Doron et al., 2018) e conduziram estudos evolutivos bacterianos de adaptação à diferentes hospedeiros (Lebreton et

al., 2017). Além disso, desde o início da era genômica, os genomas de referência individuais para cada espécie têm sido a base para compreender a variação genética. A pangenômica surge como uma ferramenta que complementa informações genéticas ausentes em um único genoma de referência, revelando variações genéticas ocultas e demonstrando a verdadeira diversidade genética em nível de espécie (Eisenstein, 2023).

Em resumo, os pangenomas representam uma abordagem inovadora na genômica comparativa, abrindo portas para uma compreensão mais profunda da diversidade genética, evolução e adaptação de espécies, contribuindo significativamente para o avanço da pesquisa científica.

### 3. OBJETIVOS

#### 3.1 OBJETIVO GERAL

Analisar a diversidade genômica e a taxonomia dos vírus do grupo *extended Asfarviridae*, visando aprimorar a compreensão de sua biologia e evolução.

#### 3.2 OBJETIVOS ESPECÍFICOS

- Investigar as diferenças genômicas entre os membros do grupo *extended Asfarviridae*, destacando variações no tamanho do genoma, conteúdo G+C, e número e função das proteínas codificadas.
- Construir o pangenoma desses vírus, identificando os genes *core* e determinar sua natureza aberta ou fechada.
- Analisar a taxonomia desses vírus, utilizando métodos filogenéticos e métricas de identidade genética para determinar suas relações evolutivas, e investigar a adequação das classificações taxonômicas (família, gênero e espécie) existentes para esse grupo viral.

## 4. MATERIAL E MÉTODOS

### 4.1 OBTENÇÃO DOS GENOMAS E ANOTAÇÃO GÊNICA.

Genomas de isolados virais de membros de *extended Asfarviridae* disponibilizados publicamente foram baixados por busca ativa no banco de dados do NCBI. Apenas genomas completos foram baixados. Uma re-anotação gênica foi realizada para cada um desses genomas, objetivando a eliminação do viés do método de anotação e homogeneização metodológica. Os genomas foram submetidos ao GeneMarkS v4.28 (Besemer et al., 2001) para predição das regiões de codificação. Essa ferramenta possui os melhores *benchmarks* para predição gênica em vírus que infectam eucariotos (González-Tortuero et al., 2021).

### 4.2 AGRUPAMENTO DAS SEQUÊNCIAS PROTEICAS

Afim de estipular as famílias de genes ortólogos entre as diferentes espécies virais, as sequências aminoacídicas de proteínas preditas foram agrupadas em *clusters* de elementos possivelmente ortólogos com o auxílio da ferramenta Proteinortho v6.0.14 (Klemm et al., 2023; Lechner et al., 2011). Os parâmetros de 60% de cobertura e 20% de identidade para agrupamento das proteínas foram adotados com base em Andreani et al., 2017.

O pangenoma foi categorizado da seguinte maneira de acordo com o nível de compartilhamento dos genes:

<b>Grupo</b>	Genes presentes em
<b>Core</b>	100% dos genomas
<b>Soft core</b>	90% dos genomas
<b>Shell</b>	$\geq 3$ genomas, $< 90\%$ genomas
<b>Cloud</b>	$\leq 2$ genomas

Com o objetivo de compreender a dinâmica de compartilhamento de genes ortólogos entre as diversas espécies virais, foi desenvolvida uma matriz numérica que representa a interseção entre os genes compartilhados nos diferentes genomas virais. Os valores foram normalizados pela quantidade total de proteínas de cada genoma e, em seguida, visualizados com um *heatmap* produzido no R com o pacote pheatmap v1.0.12 (Kolde, 2019).

*Scripts* em Python foram aplicados para conversão da tabela gerada pelo Proteinortho em uma matriz binária de presença e ausência de genes e a criação de uma curva de acumulação do pangenoma total e genes *core* (github.com/isabelschober/proteinortho\_downstream). A lei de Heaps foi aplicada à matriz de presença e ausência de genes para o cálculo do valor de  $\alpha$  com o intuito de prever se o pangenoma é de natureza aberta ou fechada. O cálculo foi

realizado no R com o pacote Micropan (Snipen & Liland, 2015) no modo “*random*” com 1.000 permutações.

A matriz de presença e ausência de genes, excluindo o grupo dos genes *core*, também foi utilizada para análise de componentes principais (PCA) associados à composição gênica assessória. A função `prcomp` do R foi utilizada para o cálculo dos componentes principais e variância explicada, combinando-os com uma tabela de metadados que inclui o gênero/grupo dos vírus, o hospedeiro, a arquitetura do genoma e tamanho do genoma. O tamanho do genoma foi categorizado em pequeno, médio e grande, com base nos quartis calculados com a função `quantile`, no R. Os gráficos de PCA foram gerados com o `ggplot2` (Wickham et al., 2024).

#### 4.3 DESCRIÇÃO DAS CARACTERÍSTICAS GENÔMICAS

Os valores referentes ao tamanho do genoma (kb), ao número de proteínas preditas e ao conteúdo G+C de cada vírus foram utilizados para a construção de *boxplots* no R com o pacote `ggplot2` (Wickham et al., 2024). Adicionalmente, para verificar possíveis diferenças nas médias entre os pares de grupos, foi conduzido o teste de Mann-Whitney, considerando um nível de significância de  $p < 0,05$ .

O tamanho das proteínas, expresso em número de aminoácidos, foi calculado utilizando o módulo `ProtParam` do Biopython (Cock et al., 2009). As proteínas foram então agrupadas de acordo com seu grupo pangenômico, e o teste de Mann-Whitney foi novamente aplicado para investigar eventuais diferenças estatisticamente significativas ( $p < 0,05$ ) entre os diferentes grupos. Para ilustrar essa análise, *Violin plots* foram gerados com o `ggplot2` (Wickham et al., 2024).

#### 4.4 ANÁLISE FILOGENÉTICA COM BASE NOS GENES *CORE*.

As sequências de aminoácidos correspondentes a cada um dos genes *core* encontrados foram alinhados utilizando o MAFFT v7 (Katoh et al., 2002), através da estratégia L-INS-i. Os alinhamentos múltiplos resultantes foram então concatenados com auxílio do script Python `catfasta2phym.py` (disponível em [github.com/nylander/catfasta2phym](https://github.com/nylander/catfasta2phym)). Posteriormente, o alinhamento resultante foi refinado utilizando o Gblocks (Castresana, 2000) para eliminar regiões fracamente alinhadas.

O alinhamento múltiplo final dos genes *core* foi submetido ao FastTree v2.1 (Price et al., 2010) para construção da árvore de Máxima Verossimilhança aplicando o modelo LG, sugerido pela ferramenta ModelFinder (Kalyaanamoorthy et al., 2017) como melhor modelo de evolução de sequências para esses dados. Valores de suporte de ramo (Shimodaira-Hasegawa

(SH)) foram calculados. A árvore filogenética gerada foi visualizada no FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree>) e editada no Inkscape v1.2.1 (2022).

#### 4.5 ANÁLISE FUNCIONAL DAS PROTEÍNAS

Para prever as funções das proteínas agrupadas, as sequências de aminoácidos em seus diferentes grupos pangenômicos foram submetidas ao eggNOG-mapper v2.1.12 (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019). Essa ferramenta realiza anotações funcionais por meio da busca de ortologia em um banco de dados que contém *clusters* de proteínas pré-calculados, os quais são comparados com diversos bancos de dados de anotação funcional, incluindo COG (*Clusters of Orthologous Groups*), KEGG (*Kyoto Encyclopedia of Genes and Genomes*) e Pfam. Para essa análise, foi gerado um gráfico de barras empilhadas no ggplot2 (Wickham et al., 2024).

Além disso, para o grupo dos genes *core*, foi conduzida uma busca funcional manual através da comparação de similaridade com BLASTp contra o banco de proteínas do nr. Foram adotados parâmetros mínimos de *e-value* mínimo de  $1e^{-3}$  e cobertura mínima de 30%. Os cinco principais hits foram considerados para cada proteína representativa. As proteínas hipotéticas, sem anotação funcional definida tanto pelo eggNOG-mapper quanto pelo BLASTp, foram submetidas a uma terceira abordagem, na tentativa de inferir funcionalidade via estrutura. Elas tiveram sua estrutura tridimensional modelada pelo ColabFold v1.5.5 (Mirdita et al., 2022), uma implementação do AlphaFold2 no google Colab, com subsequente busca de homologia estrutural no Foldseek (van Kempen et al., 2024) se utilizando do modelo *rank* 1 produzido pelo ColabFold. Apenas as estruturas com *e-value*  $< 1e^{-05}$ , probabilidade acima de 0.8, e pertencentes ao AFDB (AlphaFold Database) e PDB100 foram consideradas.

#### 4.6 CÁLCULO DE AAI e ANI

Com o propósito de estimar os níveis de similaridade entre os genomas em termos de composição nucleotídica e aminoácídica, os genomas e as sequências de aminoácidos, respectivamente, foram submetidos ao ANI/AAI *calculator* (Rodriguez-R & Konstantinidis, 2016) a fim de calcular os valores de *Average Aminoacid Identity* (AAI) e *Average Nucleotide Identity* (ANI). Essa ferramenta estima a identidade média de nucleotídeos e aminoácidos através de BLAST usando uma abordagem bidirecional onde os *best hits* (*one-way* ANI/AAI) e os *best hits* recíprocos (*two-way* AAI/ANI) entre dois conjuntos de dados genômicos de proteínas são computados.

Matrizes foram criadas com os valores par-a-par e submetidas ao R para geração de *heatmaps* com o pheatmap v1.0.12 (Kolde, 2019). Em seguida, *clusters* hierárquicos foram



calculados com a função *hclust* do R, utilizando o método de ligação completa, e os genomas foram reorganizados em dendogramas com o pacote *dendsort* (Sakai & Biederstedt, 2021). Para fins de comparação, valores médios de AAI também foram calculados para genomas de outras famílias de NCLDV (*Iridoviridae*, *Marseilleviridae*, *Ascoviridae*, *Mimiviridae*) e então plotados com *ggplot2*.

#### 4.7 ANÁLISE DE SINTENIA

A manutenção da ordem gênica é um indicativo importante de história evolutiva compartilhada. Para essa análise, as sequências das proteínas preditas foram empregadas em um BLASTp (1e-5) todos-contra-todos. A tabela resultante do BLASTp foi filtrada, selecionando hits com  $p\text{-ident} \geq 20\%$  e  $\geq 70\%$ . Em seguida, MCScanX, uma ferramenta que detecta a colinearidade de genes em pares de genomas (Y. Wang et al., 2012), foi utilizada para essa finalidade. MCScanX então utilizou as tabelas do BLASTp em conjunto com os arquivos GFFs, gerados pelo GeneMarkS contendo a localização das proteínas preditas, para avaliar a presença de blocos gênicos sintênicos entre os pares de genomas. Foram realizadas duas análises, uma para  $p\text{-ident} \geq 20\%$  e outra para  $p\text{-ident} \geq 70\%$ . Os resultados foram empregados na criação de uma matriz de sintenia, visualizada e baixada através do Synvisio (Bandi & Gutwin, 2020). Posteriormente, as duas matrizes foram combinadas e editadas no Inkscape v1.2.1 (2022).

## 5. RESULTADOS

### 5.1 VARIAÇÕES GENÔMICAS SIGNIFICATIVAS EM *EXTENDED Asfarviridae*

A busca por genomas de espécies virais de *extended Asfarviridae* no banco de dados do NCBI resultou em 39 genomas completos, sendo 18 ASFV, 16 faustovírus, 2 kaumoebavírus, 2 pacmanvírus e 1 AbALV, obtidos no formato “.fasta”. Os atributos analisados incluíram o comprimento do genoma, sua arquitetura e conteúdo G+C. Posteriormente, os genomas foram submetidos a anotação utilizando a ferramenta GeneMarkS, e a quantidade de proteínas preditas também está evidenciada na **tabela 1**.

**Tabela 1.** Lista de genomas de *extended Asfarviridae* selecionados para o estudo. Descritos estão o número de proteínas preditas na anotação, o comprimento do genoma em *base pairs* (bp), a arquitetura do genoma, o conteúdo G+C e o identificador do NCBI.

Vírus	Nº de Proteínas	Comprimento do genoma (bp)	Arq. do genoma	Conteúdo G+C (%)	NCBI ID
AbALV	355	281,224	linear	31.3	LC637659.1
ASFV OG10 Italy	163	182,906	linear	38.6	NC_044947.1
ASFV 47/ss/2008	166	184,638	linear	38.5	NC_044955.1
ASFV BA71V	148	170,101	linear	38.9	NC_001659.2
ASFV Benin 97/1	164	182,284	linear	38.6	NC_044956.1
ASFV E75	171	181,187	linear	38.7	NC_044958.1
ASFV Georgia 2007/1	172	190,584	linear	38.4	NC_044959.2
ASFV ken05/TK1	175	191,058	linear	38.3	NC_044945.1
ASFV ken06.Bus	167	184,368	linear	38.5	NC_044946.1
ASFV Kenya 1950	173	193,886	linear	38.4	NC_044944.1
ASFV L60	167	182,362	linear	38.6	NC_044941.1
ASFV Malawi 1983	174	187,612	linear	38.0	NC_044954.1
ASFV Mkuzi 1979	178	192,714	linear	38.4	NC_044953.1
ASFV NHV	158	172,051	linear	38.9	NC_044943.1
ASFV OURT 88/3	158	171,719	linear	38.9	NC_044957.1
ASFV Pretoriuskop/96/4	173	190,324	linear	38.4	AY261363.1
ASFV Tengani 62	169	185,689	linear	38.6	NC_044951.1
ASFV Warmbaths	177	190,773	linear	38.4	NC_044950.1
ASFV Warthog	173	186,528	linear	38.6	NC_044949.1
Faustovirus D3	492	455,803	circular	37.8	KU556803.1
Faustovirus D5a	489	466,051	circular	36.2	KU702950.1
Faustovirus D5b	496	464,523	circular	37.7	KU702949.1
Faustovirus D6	494	462,011	circular	37.7	KU702951.1
Faustovirus E12	489	466,265	circular	36.2	KJ614390.1
Faustovirus E23	489	465,956	circular	36.2	KU702952.1
Faustovirus E24	489	466,012	circular	36.2	KU702948.1
Faustovirus E9	521	491,024	linear	39.6	MT335755.1

Faustovirus LCD7	515	477,407	linear	39.9	MN830294.
Faustovirus liban	492	470,731	circular	36.7	MN534311.1
Faustovirus M6	505	472,803	linear	39.8	MN830295.1
Faustovirus S17	498	476,423	linear	39.6	MN830296.1
Faustovirus ST1	506	470,659	linear	36.7	LT839607.1
Faustovirus VV10	479	456,714	circular	37.7	MN956669.1
Faustovirus vv57	508	478,172	linear	39.7	MN830297.1
Faustovirus VV63	510	479,542	linear	39.7	MN830298.1
Kaumoebavirus KLCC10	513	362,586	linear	43.1	MT334784.1
Kaumoebavirus Sc	461	350,731	circular	43.7	NC_034249.1
Pacmanvirus A23	470	395,405	linear	33.6	NC_034383.1
Pacmanvirus S19	505	418,588	linear	33.2	MZ440852.1

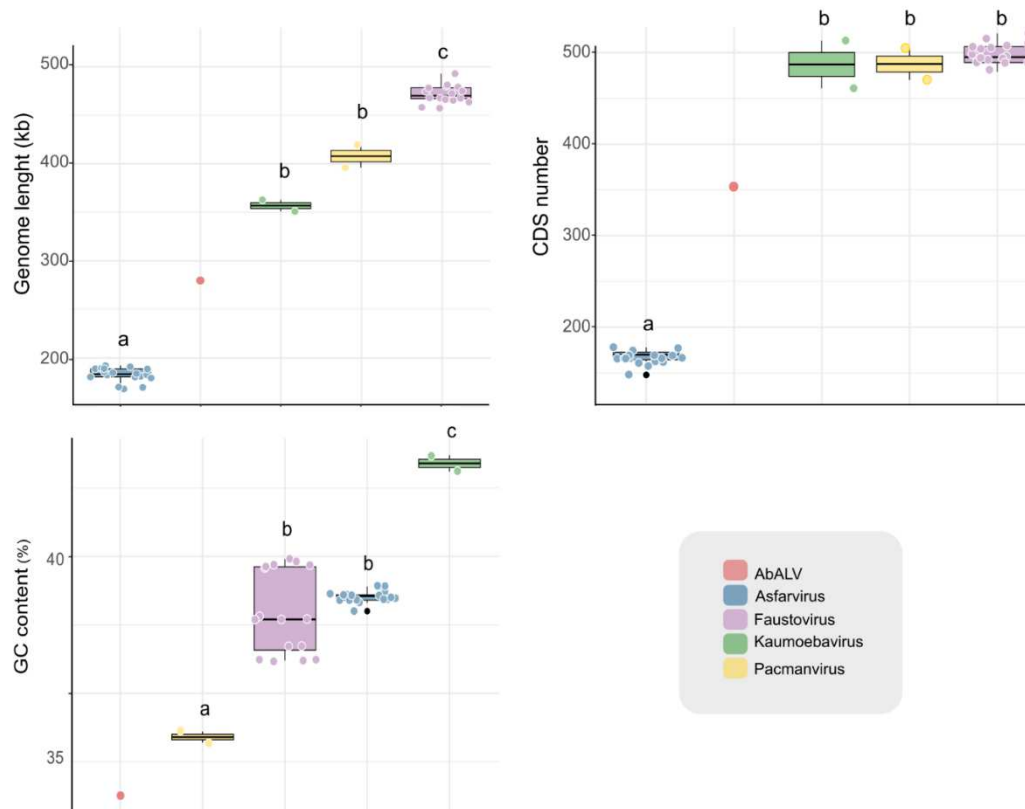
AbALV – *Abalone asfar-like virus*; ASFV – *African swine fever virus*.

Para fins de comparação, os valores de tamanho do genoma, conteúdo G+C e quantidade de proteínas preditas foram contabilizados em *boxplots* (**figura 12**). O tamanho dos genomas dentro de cada grupo revela uma distribuição centrada em torno da mediana. Ao comparar entre os grupos, observa-se uma clara tendência de aumento na seguinte ordem: ASFV > AbALV > kaumoebavírus > pacmanvírus > faustovírus (**figura 12**). Não foram encontradas diferenças estatisticamente significativas (Mann-Whitney  $p < 0,05$ ) entre os tamanhos dos genomas de pacmanvírus e kaumoebavírus, mas houveram divergências significativas durante a comparação dos outros grupos. É interessante notar a grande diferença no comprimento do genoma entre ASFV e o grupo com o maior genoma do clado, os faustovírus. Os faustovírus possuem, em média, um genoma 2,5 vezes maior que os ASFV.

Faustovírus e asfarvírus não apresentaram diferenças significativas no conteúdo G+C entre seus membros. AbALV possui o menor conteúdo G+C do grupo, sendo quase 40% menor que o maior conteúdo G+C, encontrado em Kaumoebavirus Sc. Destaca-se a separação dos faustovírus em três faixas de conteúdo G+C: Faustovírus D5a, E12, E23, E24, Liban, ST1, com média de 36.3%; Faustovírus D3, D5b, D6, VV10, com média de 37.7%; e Faustovírus E9, LCD7, M6, S17, vv57, VV63, com média de 39.5%.

No que diz respeito ao número de proteínas preditas, os ASFV continuam a apresentar os menores valores (**figura 12**), enquanto kaumoebavírus, pacmanvírus e faustovírus não demonstram diferenças estatisticamente significativas. É notável que, apesar de os genomas de kaumoebavírus e pacmanvírus serem algumas bases de pares mais curtas, eles ainda mantêm um número de proteínas próximo ao dos faustovírus, o grupo com o maior número médio de tamanho de genoma.

**Figura 12** – *Boxplots* representando a variação em tamanho do genoma, número de proteínas preditas e conteúdo G+C em 39 membros de *extended Asfarviridae* separados por grupo. Letras minúsculas diferentes representam diferença estatisticamente significativa (Mann-Whitney  $p < 0.05$ ).



AbALV – *Abalone asfar-like virus*.

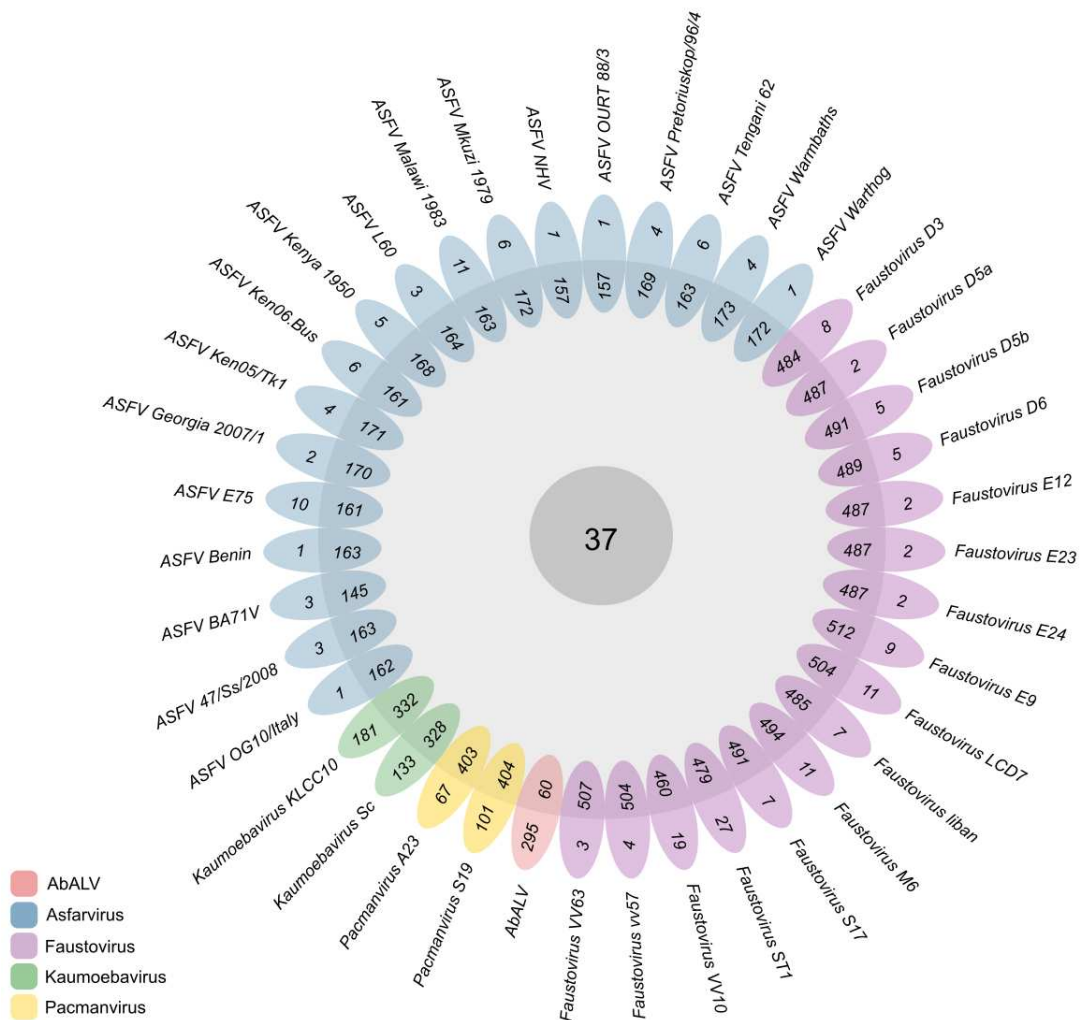
## 5.2 EXTENDED *Asfarviridae* APRESENTA UM PANGENOMA ABERTO EXTENSO E UM *CORE* LIMITADO.

O processo de agrupamento das sequências proteicas com base na similaridade resultou em um total de 2.483 clusters de proteínas ortólogas ou genes *singletons* para *extended Asfarviridae*. Apenas 37 proteínas são encontradas em todos os 39 genomas pesquisados, enquanto 973 proteínas foram consideradas *singletons* (**figura 13**), ou seja, não foram encontradas em dois ou mais genomas.

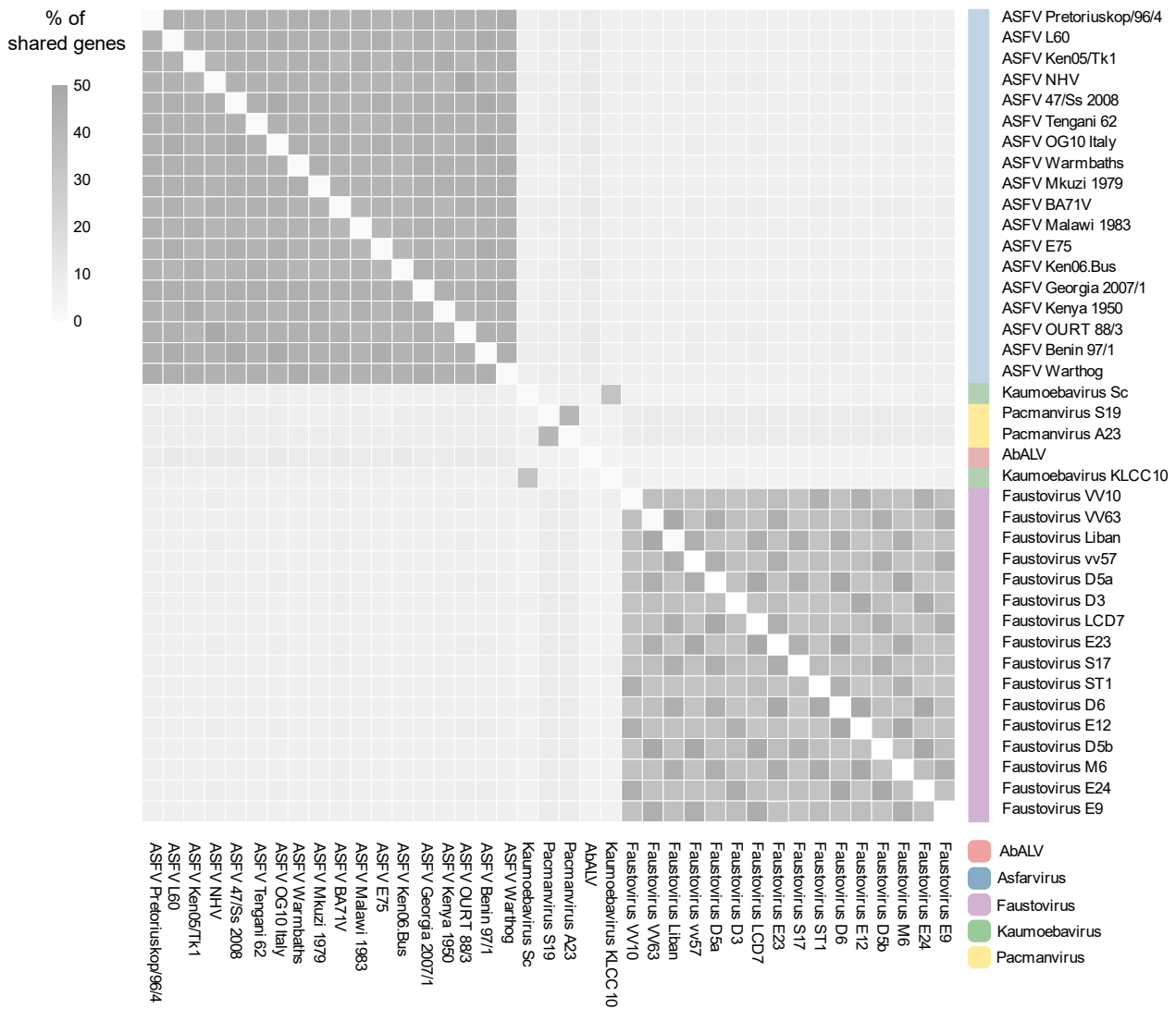
Em relação aos genes *core*, observou-se que, para cada linhagem analisada, correspondiam a aproximadamente 7~8% do total de genes nos faustovírus, kaumobavírus e pacmanvírus. No entanto, o cenário é diferente no caso do AbALV, no qual os genes *core* representam 10% do total de genes, e nos ASFV, onde essa proporção varia de 20 a 25% do genoma.

Por outro lado, considerando os genes *singletons*, a proporção varia significativamente entre as linhagens. Para os faustovírus, a variação vai de 0 a 5, enquanto para os ASFV é de 0 a 6%. O AbALV apresenta uma proporção de 83% de *singletons*, enquanto o pacmanvírus e o kaumobavírus têm proporções entre 14% e 20%, e 29% e 35%, respectivamente.

**Figura 13** – Flower Plot mostrando os genes *core*, acessórios e *singletons* de 39 membros de *extended Asfarviridae*. O flower Plot exibe o número de genes *core* (no centro), o número de genes acessórios (dentro do anel) e o número de genes *singletons* (nas pétalas). AbALV – *Abalone arfar-like virus*; ASFV – *African swine fever virus*.



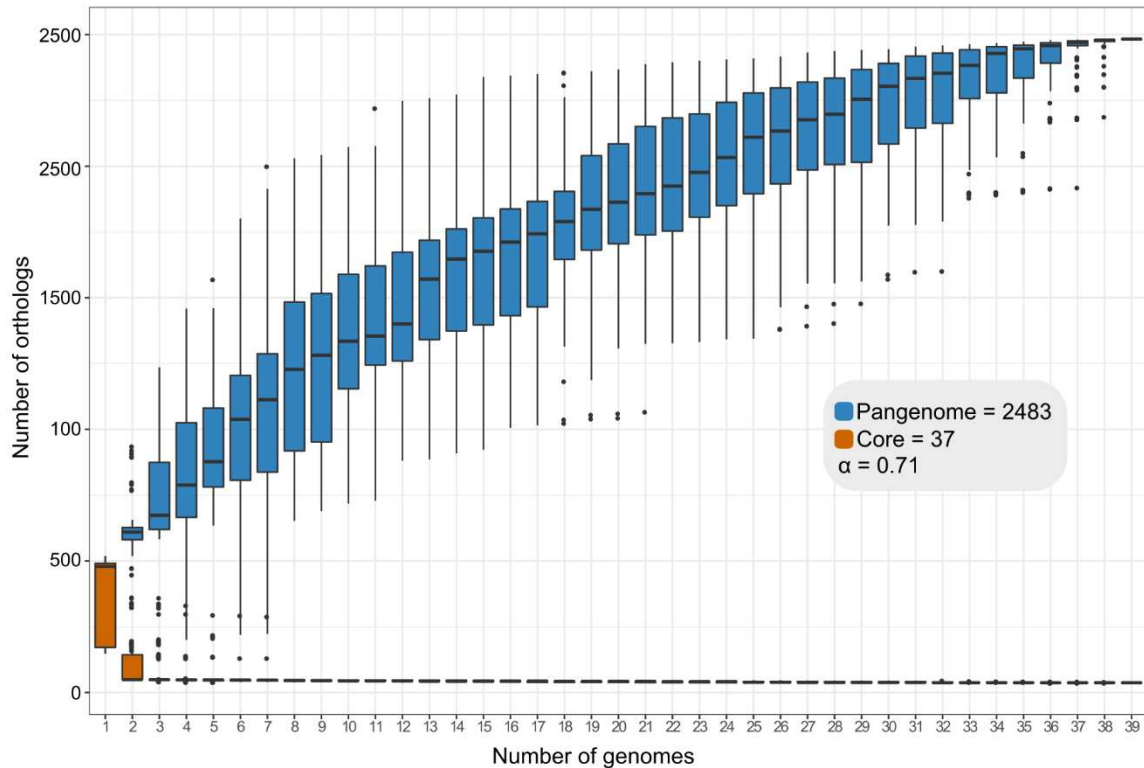
**Figura 14** – Heatmap representando a proporção de genes compartilhados entre cada par de membros de *extended Asfarviridae*. AbALV – *Abalone arfar-like virus*; ASFV – *African swine fever virus*.



A dinâmica de compartilhamento de genes neste clado revela uma distinção clara entre os grupos, conforme ilustrado na **figura 14**. Olhando a variação intragrupo, nos ASFVs, o compartilhamento de genes varia de 41% a 49%, enquanto nos faustovírus varia entre 32% e 50%. Os genomas dos dois kaumobavírus, apesar de possuírem maior proximidade genética, compartilham apenas 33% de suas proteínas, representando a maior proporção de compartilhamento entre kaumobavírus e qualquer outro genoma. Os pacmanvírus compartilham um pouco mais, cerca de 41% das proteínas. Em relação a este aspecto, o AbALV mostra maior proximidade com os ASFVs, compartilhando aproximadamente 10 a 11% das

proteínas. Todos os outros genomas, quando comparados com vírus fora de seus grupos ou gêneros, compartilham menos de 10% das proteínas.

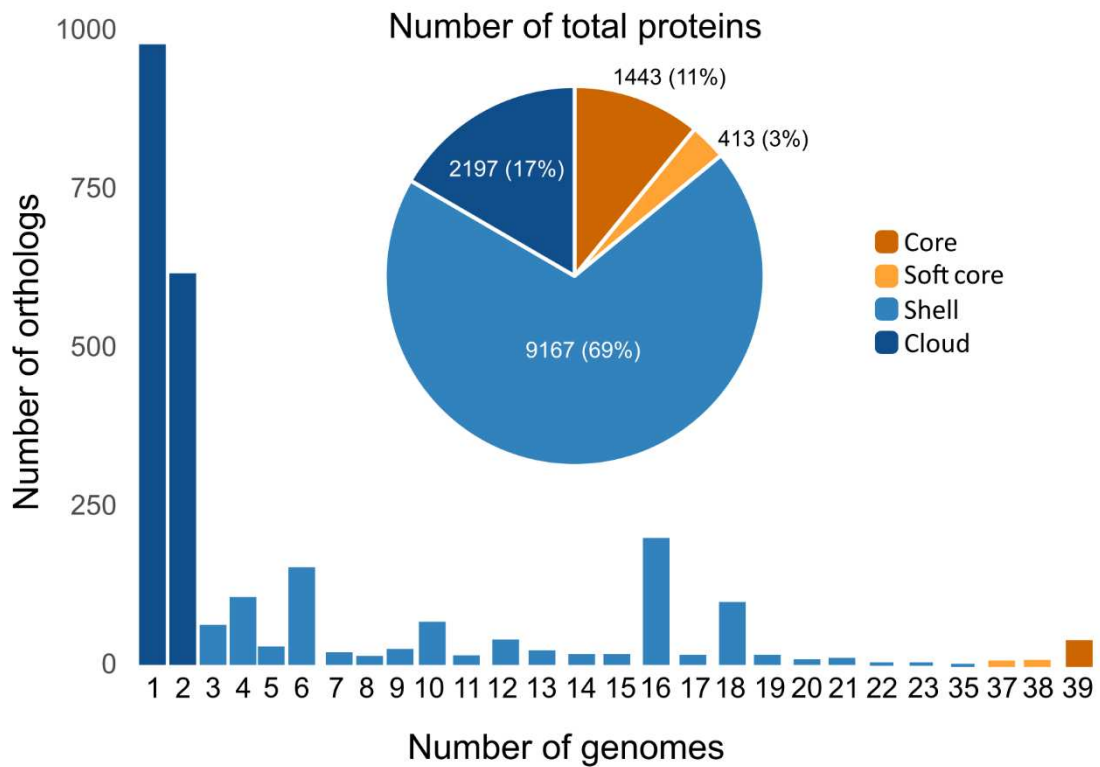
**Figura 15** – Curva de acumulação de genes para o pangenoma total e genes *core* de *extended Asfarviridae*. O valor de  $\alpha$  indica a lei de potência de Heaps para o número médio de novos genes por genoma após 1.000 permutações do pangenoma. Abaixo de 1 indica que o pangenoma está em estado "aberto" (Tettelin et al., 2005).



A inclusão sequencial de genomas representantes de *extended Asfarviridae* não evidencia claramente a formação de um platô ao final da curva, sugerindo que novas inclusões ainda podem aumentar o número de novos genes (**figura 15**). No eixo x, estão representados os 39 genomas adicionados sequencialmente, enquanto o eixo y mostra os números de ortólogos que surgem após cada adição. Após a aplicação da regressão de potência ( $n = \kappa N^{-\alpha}$ ) sobre o número total de genes em diferentes combinações de linhagens, o parâmetro  $\alpha$  foi estimado como 0,71. A interpretação desse parâmetro  $\alpha$  ( $\alpha < 1$  é aberto), juntamente com a ausência de um claro platô ao final da curva, indica que o número de genes únicos no pangenoma de *extended Asfarviridae* ainda não atingiu um patamar com o aumento do número de linhagens, sugerindo um pangenoma "aberto" para esse clado. É importante destacar que a curva de genes *core* estabiliza rapidamente. Ao contrário da curva de pangenoma, esse grupo de genes *core* se mostra conservado e com pouca variação à medida que novos genomas são incluídos.

Ao estratificar os clusters de proteínas ortólogas conforme seu grupo pangenômico, ou seja, de acordo com o nível de compartilhamento das proteínas, os padrões observados são ilustrados na **figura 16**. O histograma retrata a frequência gênica de ortólogos em cada genoma, enquanto o gráfico de pizza ilustra a proporção das proteínas em cada grupo pangenômico no pangenoma global. O histograma indica que a maioria dos genes está presente em um número reduzido de genomas (genes acessórios), enquanto poucos genes estão presentes em todos os genomas (genes *core*), sugerindo uma alta variação no conteúdo gênico dos vírus neste clado. Além disso, nota-se uma predominância das proteínas *singletons/cloud* nos clusters, embora as proteínas *shell* liderem em termos de número total de proteínas no gráfico de pizza.

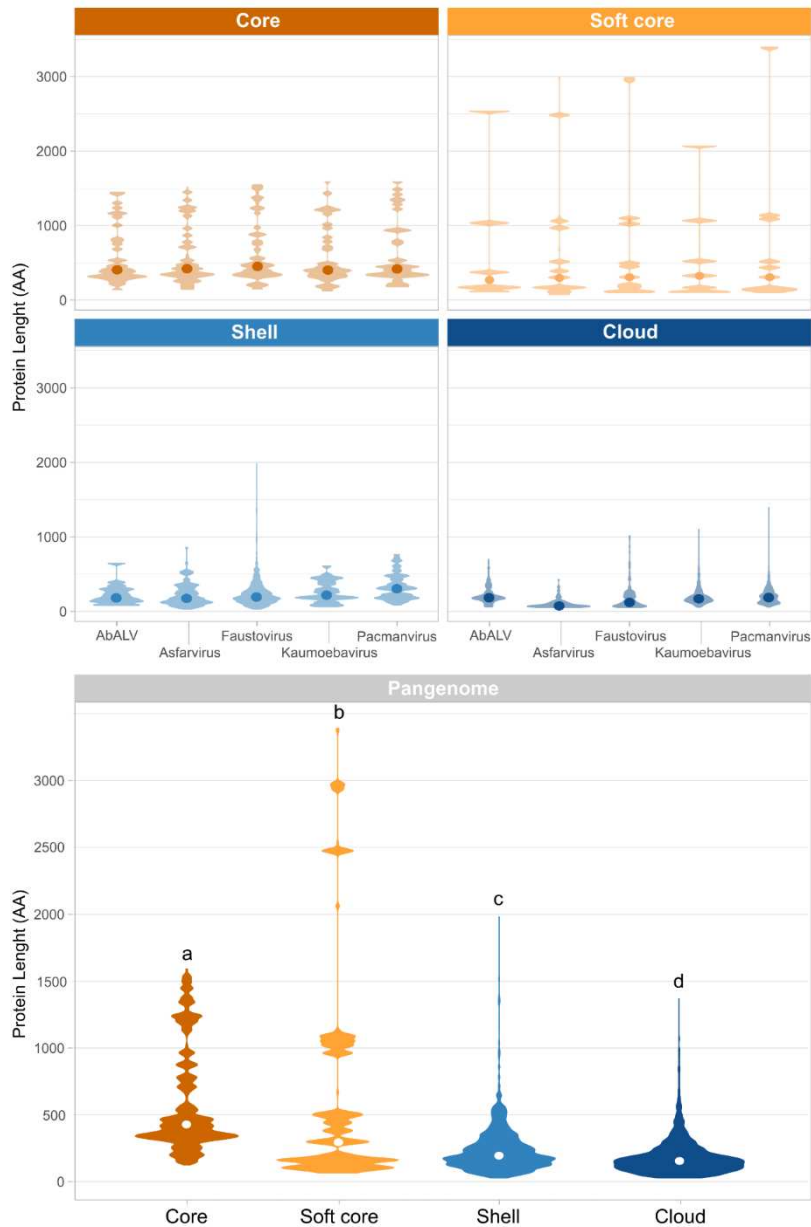
**Figura 16** – Histograma das distribuições dos genes nos grupos pangenômicos *core*, *soft core*, *shell* e *cloud*. O gráfico de pizza exibe o número total de genes com as porcentagens entre parênteses.



As proteínas desse clado apresentam variações significativas em seu tamanho quando agrupadas por grupo pangenômico (**figura 17**). Notavelmente, as proteínas do grupo *core* tendem a ser as maiores em comparação com os outros grupos de vírus, além de serem as maiores em média em relação a todas as proteínas do pangenoma. Em termos de tamanho, as proteínas do grupo *core* são seguidas pelas do *soft core*, *shell* e *cloud*, e todas mostraram diferenças estatisticamente significantes em suas médias (Mann-Whitney,  $p < 0.05$ ). Por outro lado, as proteínas do grupo *cloud* tendem a ser as menores, em todos os cenários.



**Figura 17** – Tamanho médio das proteínas preditas do pangenoma de *extended Asfarviridae* separadas por grupo pangênômico ou total. AbALV – *Abalone asfar-like virus*. Letras minúsculas diferentes representam diferença estatisticamente significativa (Mann-Whitney  $p < 0.05$ ).

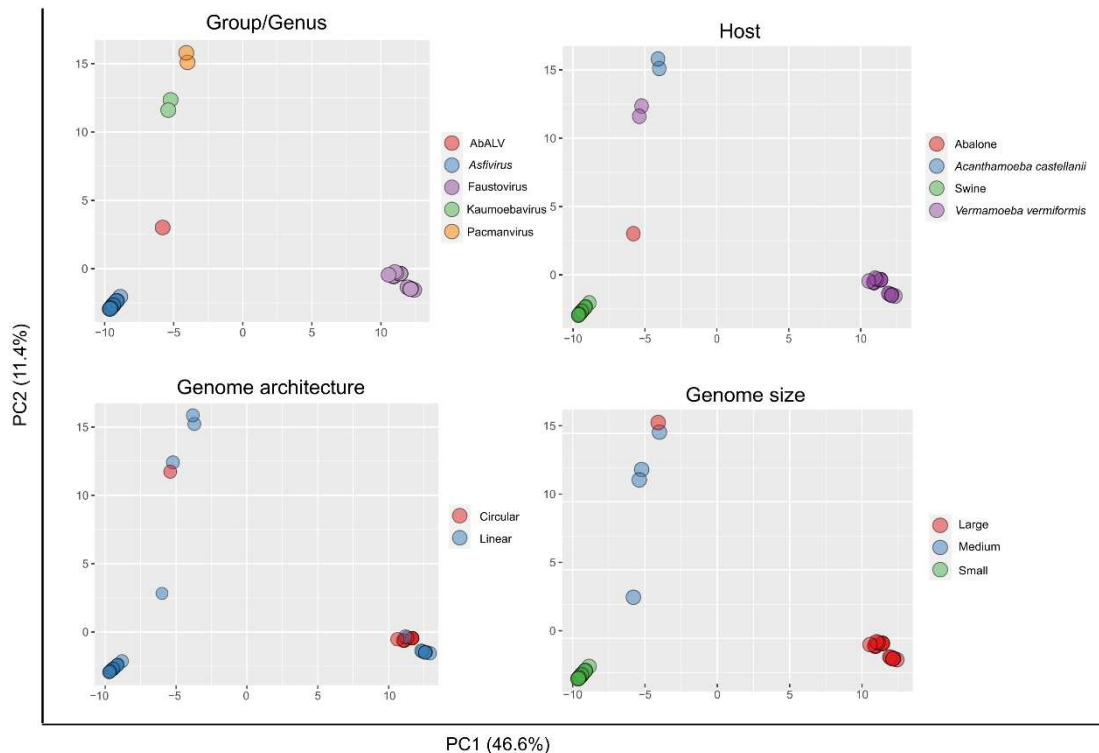


A Análise de Componentes Principais (PCA) é utilizada como uma ferramenta estatística para explorar a diversidade genômica e identificar possíveis associações entre genomas com base no conteúdo gênico e extrapolar para algum metadado associado. Nesse método, a matriz binária de presença/ausência de genes, sem o grupo de genes *core*, é transformada em um conjunto de componentes principais, os quais capturam a variação no conteúdo gênico entre os diferentes genomas. Esses componentes principais são então plotados em um espaço bidimensional, onde cada ponto representa um genoma. A posição de cada ponto

reflete seu conteúdo gênico, e a distância entre os pontos indica a divergência entre os genomas em termos de conteúdo gênico.

Os resultados da Análise de Componentes Principais (PCA), conforme mostrado na **figura 18**, revelam uma clara segregação dos genomas com base em seus genes acessórios, observada pela formação de agrupamentos principalmente ao longo dos Componentes Principais 1 (PCA1) e 2 (PCA2). Essa distribuição, juntamente com os metadados associados, sugere uma possível associação entre o conteúdo genético acessório e o grupo/gênero viral, ilustrando que a taxonomia consegue refletir o conteúdo gênico, bem como possíveis vínculos com o hospedeiro e até mesmo com o tamanho do genoma, indicando que esses fatores podem ser determinantes significativos do conteúdo gênico. Essa associação é principalmente capturada pelo Componente Principal 1. Não parece haver influência da arquitetura do genoma sobre o conteúdo gênico acessório

**Figura 18**– Análise de Componentes Principais (PCA) baseada na presença/ausência de genes acessórios para 39 genomas de *extended Asfarviridae*. Os indivíduos são plotados de acordo com seus escores nos dois maiores componentes principais. As variâncias explicadas para cada componente principal se encontram entre parênteses.



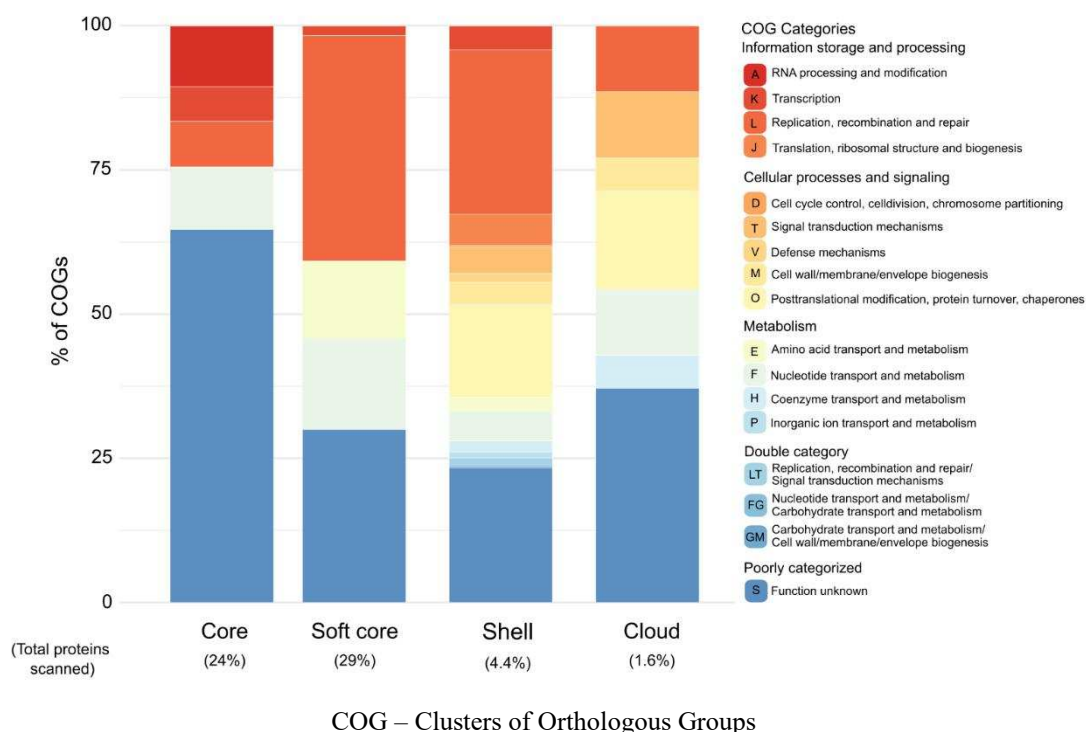
Observa-se uma distinção (PC1) entre os vírus que infectam eucariotos unicelulares (faustovírus, kaumobavírus e pacmanvírus) e aqueles que afetam eucariotos multicelulares (ASFV e AbALV), indicando uma possível influência do fator uni ou multicelularidade na

composição genômica. Além disso, a proximidade entre kaumobavírus e pacmanvírus nos painéis sugere um compartilhamento de genes acessórios entre esses vírus independente do hospedeiro, já que eles foram isolados de diferentes hospedeiros. O papel do hospedeiro parece ser menos determinante, pois, caso fosse, kaumobavírus estaria mais próximo de faustovírus, ambos compartilhando o mesmo hospedeiro. Outro aspecto relevante é a proximidade genética entre ASFV e AbALV, sugerindo uma possível relação evolutiva ou funcional entre esses dois grupos virais, talvez relacionada à adaptação à multicelularidade.

### 5.3 ANOTAÇÃO FUNCIONAL EM EXTENDED *Asfarviridae* APRESENTA DESAFIOS

Apenas 4,67% dos *clusters* de proteínas ortólogas apresentaram correspondência com o banco de dados do COG, sendo 24% do *core*, 29% do *soft core*, 4,4% do *shell* e 1,6% do *cloud* (figura 19).

**Figura 19** – Classificação funcional dos *clusters* de proteínas ortólogas do pangenoma de *extended Asfarviridae* com base no banco de dados do *Clusters of Orthologous Groups* (COG). Todas as proteínas preditas foram alinhadas ao banco de dados COG a fim de prever funções. As letras maiúsculas no eixo Y indicam as categorias COG conforme legenda. A porcentagem sinalizada abaixo do grupo pangenômico indica a proporção de transcritos classificados para a totalidade de cada grupo.



Quando se consideram todos os grupos pangenômicos, a categoria mais comum de acordo com o COG é a "S", que corresponde a proteínas sem função conhecida (figura 19). No grupo *soft core*, a categoria mais frequente é a "L", relacionada a replicação, recombinação e

reparo, indicando a manutenção da importância desse processo mesmo em um grupo menos restrito, seguida por "S", "F": Metabolismo e transporte de nucleotídeos, e "E": Transporte e metabolismo de aminoácidos. O grupo *shell* foi o grupo com a maior diversidade de funções, as categorias mais presentes são "L", ainda presente mas em menor proporção que o *soft core*, "S" e "O": Modificação pós-traducional, reciclagem de proteínas e chaperonas. Já no grupo *cloud*, predominam as categorias "S", "O", "T" Mecanismos de transdução de sinal, e "L".

O grupo *core* apresentou uma taxa de anotação significativa, atingindo 94,6% quando considerados os três métodos de anotação funcional adotados (banco de dados do COG, BLASTp e Alphafold + FoldSeek) (**tabela 2**).

**Tabela 2.** Genes do grupo *core* do pangenoma de *extended Asfarviridae* e suas funções previstas por diferentes metodologias. Marcados em azul claro estão os únicos genes que não são encontrados no genoma *core* de NCLDV.

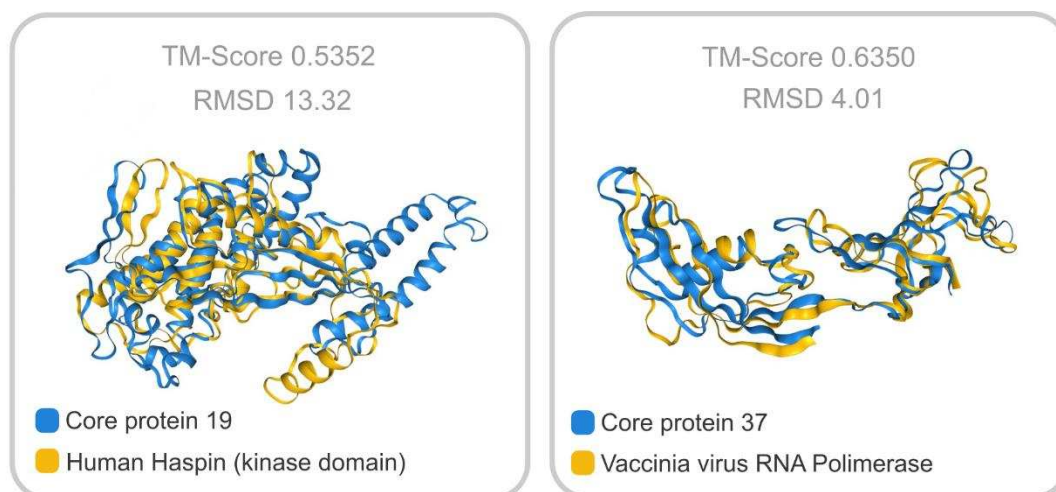
Core gene	eggNOG-mapper	BLASTp	Alphafold + Foldseek	Function
1	N-methyltransferase activity	DNA helicase	-	DNA replication, recombination and repair
2	DNA-directed RNA polymerases II	DNA-directed RNA polymerases II	-	Transcription and RNA processing
3	-	Yqaj recombinase	-	DNA replication, recombination and repair
4	Ulp1 protease, C-terminal catalytic domain	Cysteine protease	-	Other metabolic functions
5	DNA gyrase B	DNA gyrase B	-	DNA replication, recombination and repair
6	-	RNA III	-	Transcription and RNA processing
7	-	pH339R ASFV protein	-	Transcription and RNA processing
8	N-methyltransferase	Helicase	-	DNA replication, recombination and repair
9	-	Proliferating cell nuclear antigen	-	DNA replication, recombination and repair
10	-	Transcription factor S-II	-	Transcription and RNA processing
11	Ribonucleotide reductase, small chain	Ribonucleotide reductase, small chain	-	Nucleotide metabolism
12	Ribonucleotide reductase, barrel domain	Ribonucleotide reductase, barrel domain	-	Nucleotide metabolism

13	-	pK421R ASFV protein	-	Virion structure
14	RNA polymerase beta subunit	RNA polymerase beta subunit	-	Transcription and RNA processing
15	FtsJ-like methyltransferase	FtsJ-like methyltransferase	-	Other metabolic functions
16	-	ERCC4 domain-containing protein	-	DNA replication, recombination and repair
17	-	Hypothetical protein	Uncharacterized	-
18	-	<b>RNA ligase</b>	-	<b>Transcription and RNA processing</b>
19	-	Hypothetical protein	Protein kinase domain-containing protein	Other metabolic functions
20	Polynucleotide adenylyltransferase	Polynucleotide adenylyltransferase	-	Transcription and RNA processing
21	-	Putative TFIIB	-	Transcription and RNA processing
22	-	DNA-directed RNA Polymerase sub 6	-	Transcription and RNA processing
23	N-methyltransferase activity	Helicase	-	DNA replication, recombination and repair
24	-	Hypothetical protein	Uncharacterized	-
25	-	DNA packaging ATPase	-	Virion maturation
26	-	Putative MCP chaperone	-	Virion structure
27	-	VLTF3 late transcription factor	-	Transcription and RNA processing
28	-	VLTF2-like transcription factor	-	Transcription and RNA processing
29	-	TATA box binding protein	-	Transcription and RNA processing
30	-	Early transcription factor	-	Transcription and RNA processing
31	DNA polymerase family B	DNA polymerase family B	-	DNA replication, recombination and repair
32	-	pCP530R	-	Virion structure

33	RNA polymerase Rpb1, domain 3	RNA polymerase Rpb1, domain 3	-	Transcription and RNA processing
34	DNA ligase (ATP) activity	DNA ligase (ATP) activity	-	DNA replication, recombination and repair
35	mRNA capping enzyme	mRNA capping enzyme	-	Transcription and RNA processing
36	Phosphatase activity	mRNA decapping protein	-	Transcription and RNA processing
37	-	Hypothetical protein	RNA pol 2 (18kDa subunit)	Transcription and RNA processing

Das 21 proteínas do grupo *core* que não foram identificadas pelo banco de dados do COG, sua anotação foi complementada através de uma busca manual no BLASTp contra o banco de dados nr, a qual se mostrou bastante exitosa. Apesar dessa abordagem, quatro proteínas ainda permaneceram sem correspondência com proteínas anotadas e com função conhecida, um desafio comumente encontrado em vírus gigantes. Essas proteínas foram então submetidas a uma terceira análise, que envolveu a modelagem tridimensional de suas estruturas, seguida pela busca de homologia estrutural em bancos de dados de estruturas proteicas. Para duas das quatro proteínas hipotéticas, foram encontradas estruturas homólogas. A proteína *core* 19 possui uma estrutura homóloga a um domínio quinase enquanto que a proteína *core* 37 possui uma estrutura homóloga a uma subunidade (18kDa) da RNA polimerase II de Vaccinia vírus (**figura 20**).

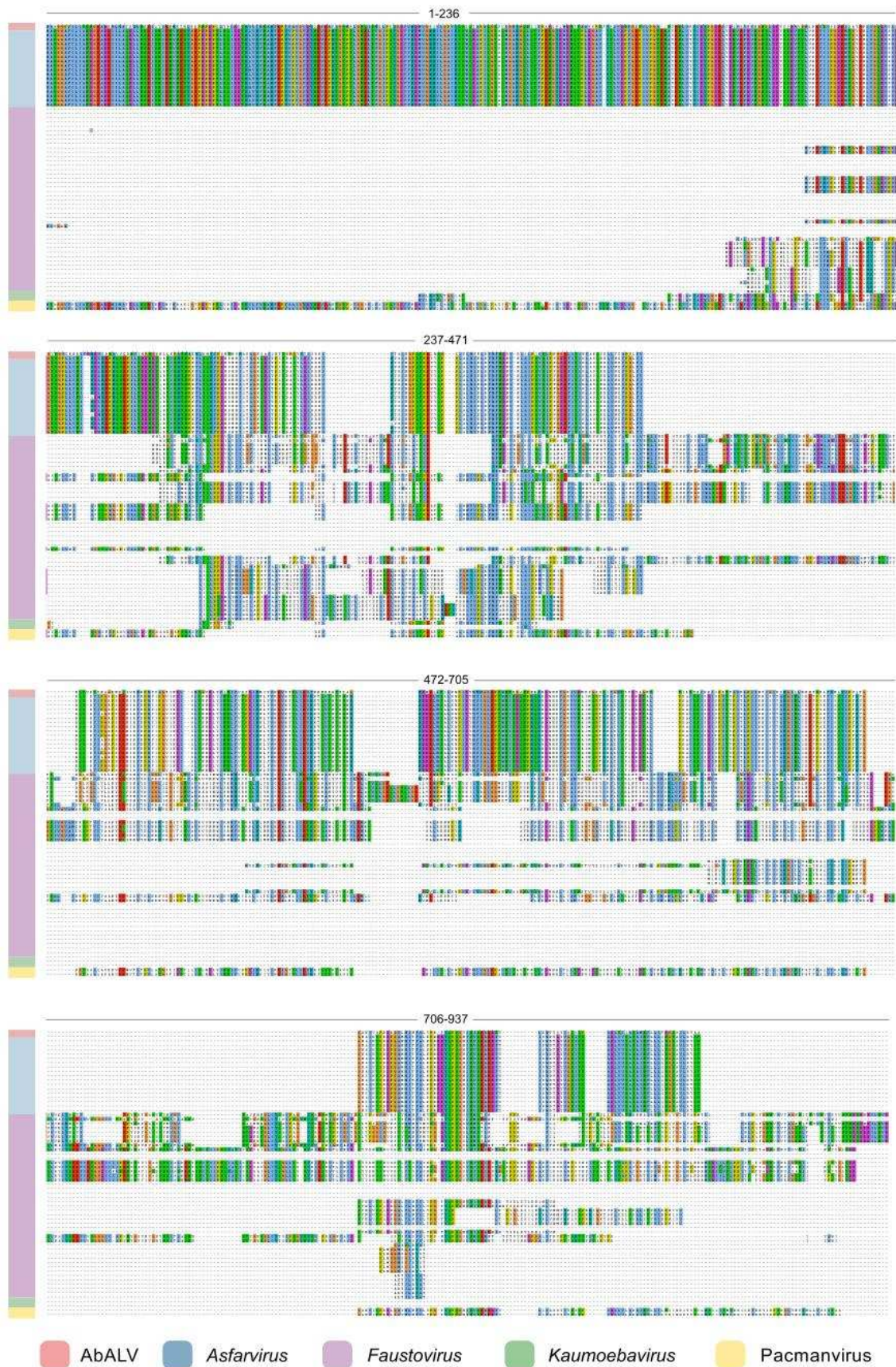
**Figura 20** – Alinhamento de estruturas tridimensionais de proteínas hipotéticas presentes no genoma *core* de *extended Asfarviridae* com estruturas homólogas.  
RMSD – Root Mean Square Deviation



Ao analisar os resultados das três etapas de anotação para o grupo *core*, observa-se que as proteínas estão envolvidas predominantemente em processos bioquímicos relacionados ao dogma central (replicação, transcrição) e, em menor número, metabolismo e estrutura viral. Comparando esse resultado com o grupo de 47 genes *core* ancestrais dos NCLDVs (Yutin et al., 2009), constata-se que 78% dos genes *core* de *extended Asfarviridae* estão presentes nesse conjunto de genes *core*.

Observa-se a ausência da proteína estrutural *Major Capsid Protein* (MCP) nos resultados obtidos. Essa proteína desempenha um papel fundamental na formação do capsídeo viral e é comumente encontrada em todos os genomas virais de espécimes do reino *Bamfordvirae*, ao qual pertencem os NCLDV. No entanto, sabe-se que, nos Faustovírus, esse gene está fragmentado ao longo do genoma (Cherif Louazani et al., 2018), o que pode ter afetado a etapa de clusterização. Portanto, foi decidido localizar os fragmentos gênicos correspondentes ao *MCP* nos Faustovírus para validar sua presença nos genomas. Isso foi realizado por meio de uma busca de similaridade utilizando o BLASTp. Como resultado, foram identificados 43 transcritos correspondentes ao *MCP* nos Faustovírus, confirmando sua presença, com uma média de 2,6 transcritos para cada genoma do grupo. Os genes foram então alinhados com todos os outros MCPs dos outros membros do grupo utilizando o MAFFT v7, estratégia L-INS-i, e o alinhamento múltiplo resultante foi visualizado no AliView (Larsson, 2014) (**figura 21**).

**Figura 21** – Alinhamento de múltiplas seqüências de amino ácidos do gene *Major Capsid Protein* (MCP) presente em 39 genomas de *extended Asfarviridae*. Genomas de faustovírus possuem 2,6 seqüências por genoma em média.

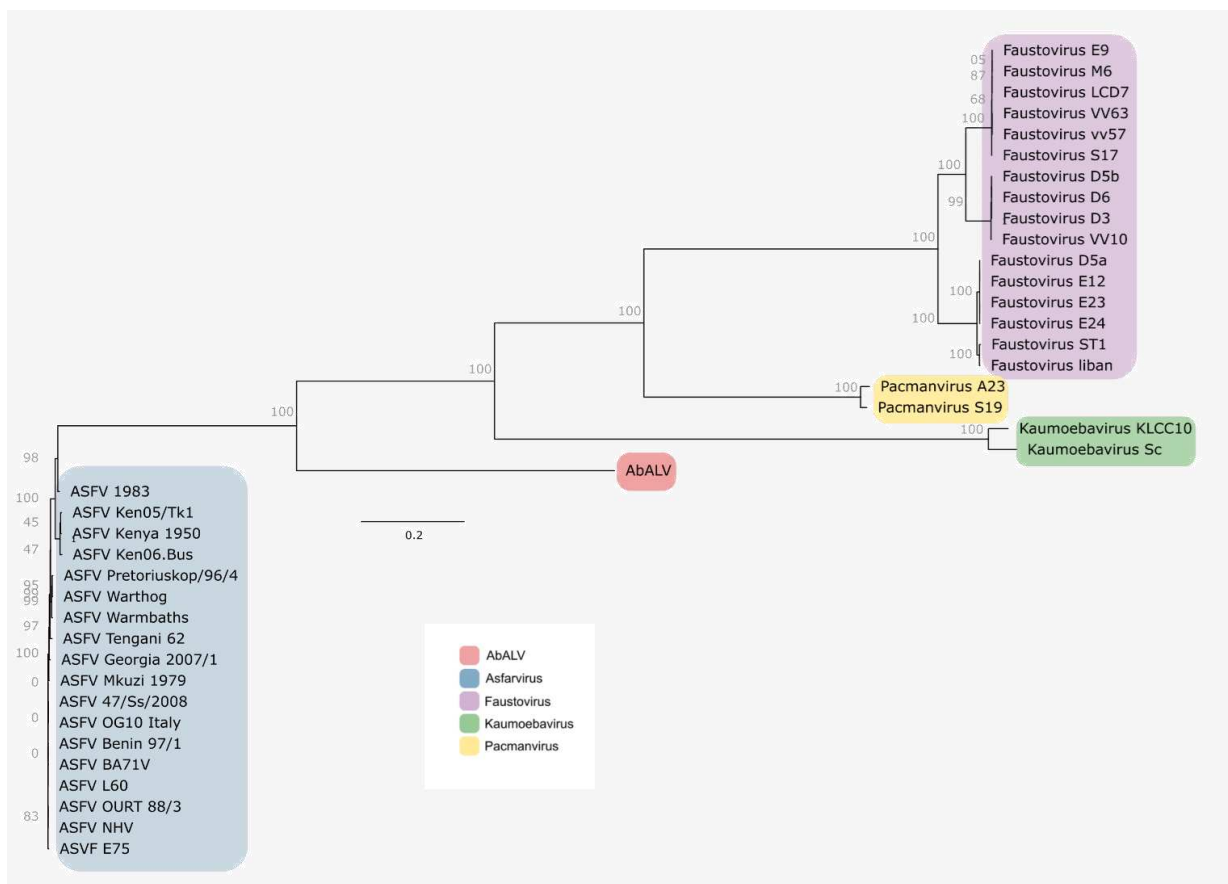




#### 5.4 FILOGENIA DOS GENES *CORE* EVIDENCIA A DISTÂNCIA EVOLUTIVA ENTRE OS MEMBROS DE *extended Asfarviridae*

Objetivando investigar as relações entre os membros de *extended Asfarviridae* e comparar as árvores filogenéticas dos genes *core* e de genes individuais (como a DNA polimerase), uma árvore filogenética de máxima verossimilhança foi construída com os 37 genes *core* concatenados. Em geral, a topologia da árvore baseada na DNA polimerase foi mantida (**figura 22**). Houve um suporte baixo para alguns dos nós mais externos em faustovírus e ASFV, mas, em geral, a análise dos valores de suporte de ramo indicou um suporte forte. É notável a distância filogenética entre os membros desse grupo, com destaque para o isolamento de ASFV.

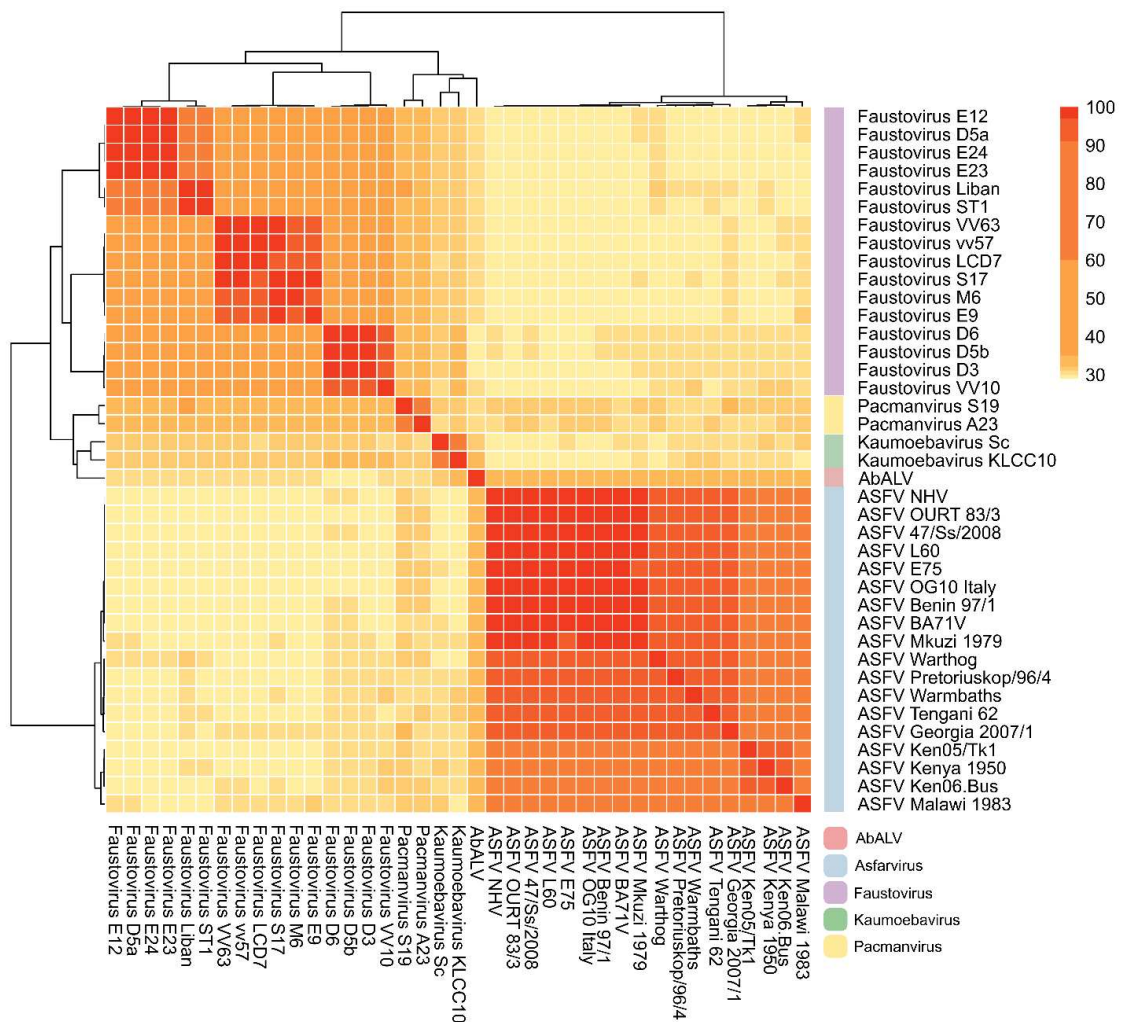
**Figura 22** – Árvore filogenética de Máxima Verossimilhança com base em 37 genes *core* concatenados de *extended Asfarviridae*. Valores de suporte de ramo são baseados em SH-aLRT



#### 5.4 OS GRUPOS DE *extended Asfarviridae* APRESENTAM ASSINATURAS PARTICULARES DE AMINOÁCIDOS E NUCLEOTÍDEOS

As análises de identidade média de aminoácidos (AAI) e identidade média de nucleotídeos (ANI) foram conduzidas com o propósito de comparar a distância genômica entre os diferentes espécimes virais em duas perspectivas distintas. Essas análises são métodos comuns empregados para determinar a similaridade entre genomas bacterianos e virais. O AAI compara a similaridade entre as sequências de aminoácidos das proteínas codificadas pelos genes dos genomas, enquanto o ANI avalia a similaridade entre as sequências de nucleotídeos dos genomas em questão. Essas análises são fundamentais para compreender a relação evolutiva e a diversidade genética entre os organismos em estudo. A matriz geral de AAI evidencia claramente a proximidade e distância entre os grupos (**figura 23**).

**Figura 23** – Heatmap de valores de identidade média de aminoácidos (AAI) par-a-par entre 39 genomas de *extended Asfarviridae*. Barra de cores é segmentada pelos quartis dos valores de AAI. AbALV – *Abalone asfar-like virus*; ASFV – *African Swine Fever Virus*.



A clusterização de dados hierárquicos na matriz do *heatmap* e a formação do dendrograma são técnicas utilizadas para identificar padrões de similaridade entre os elementos de um conjunto de dados. No contexto deste estudo, essas técnicas foram aplicadas para analisar a proximidade genética, em termos aminoácidos, entre diferentes linhagens de vírus. O método de ligação completa, utilizado neste caso, avalia a similaridade entre todos os pares de observações e, com base nisso, agrupa os elementos em *clusters* de acordo com sua distância genética. O dendrograma resultante permite visualizar a relação de parentesco entre as linhagens, mostrando quais estão mais próximas umas das outras e quais formam grupos distintos.

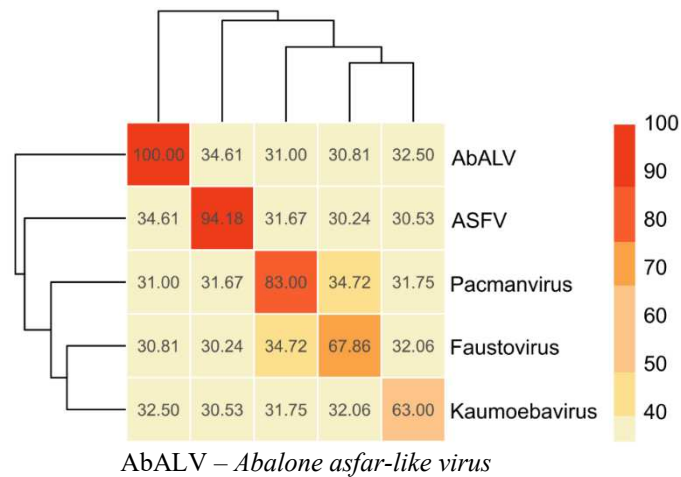
Os ASFV compartilham, internamente, um índice de AAI muito alto, acima de 95%, enquanto os valores de AAI entre os ASFV e outros grupos são significativamente mais baixos, por volta de 30%. Ainda assim, de acordo com o dendrograma, há separação de grupos. O AbALV é o que mais se aproxima dos ASFV em termos de AAI compartilhando até 32% de identidade.

Para o grupo dos faustovírus, é evidente a formação de quatro grupos com altos índices de AAI (acima de 95%), tanto visualmente no *heatmap* quanto pelo agrupamento do dendrograma: 1- Faustovírus E12, D5a, E23, E24; 2- Faustovírus ST1 e Liban; 3- Faustovírus E9, M6, S17, LCD7, vv57, VV63; e 4- Faustovírus D6, D5b, D3, VV10. Essa mesma separação surge na análise de conteúdo G+C, com a única diferença de que, na análise de G+C, os grupos 1 e 2 estão unidos. A comparação de AAI entre membros de grupos distintos resulta em uma queda no índice AAI, para cerca de 50 a 60%. Assim como os ASFV, o índice de AAI dos faustovírus com outros grupos é baixo, em torno de 30%.

Pacmanvírus e kaumobavírus mostraram alta afinidade em AAI somente intragrupo e, mesmo assim, não é tão alta como intra-linhagens de faustovírus e intra-ASFV. Os dois kaumobavírus exibem índice de AAI somente de 63% e entre os dois Pacmanvírus é um pouco mais alto, de 83% (**figura 24**). Quando esses dois grupos são comparados com ASFV, faustovírus e AbALV, os índices de AAI caem ainda mais entre 29 e 35%.

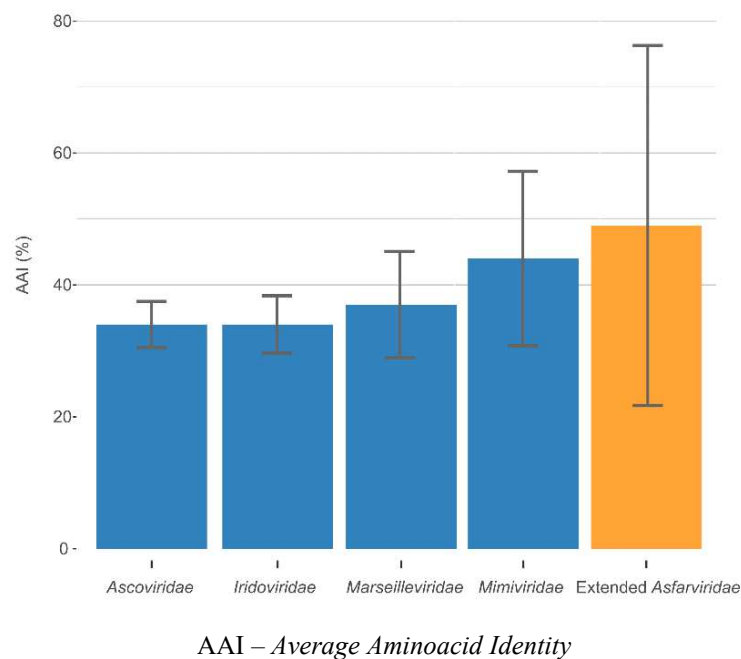
Ao considerar o valor de AAI calculado por grupo viral (**figura 24**), nota-se uma maior proximidade entre faustovírus, kaumobavírus e pacmanvírus em comparação com ASFV e seu grupo mais próximo, o AbALV. Destaca-se também o baixo índice intra-kaumobavírus (63%) e intra-faustovírus (67,86%).

**Figura 24** – Heatmap de valores de identidade média de aminoácidos (AAI) par-a-par entre todos os genomas agrupados em seus respectivos grupos. AbALV – *Abalone asfar-like virus*; ASFV – *African Swine Fever Virus*.



A média de AAI calculada para todos os 39 membros de *extended Asfarviridae* resultou em 49% (**figura 25**), um valor superior aos calculados para *Ascoviridae*, *Iridoviridae*, *Marseilleviridae* e *Mimiviridae* para fins de comparação. No entanto, há uma maior variação nos dados, como evidenciado pela barra de desvio padrão.

**Figura 25** – Médias de valores par-a-par de identidade de aminoácidos (AAI) entre membros de cada família de NCLDVs. Barras de erro indicam desvio padrão da média.



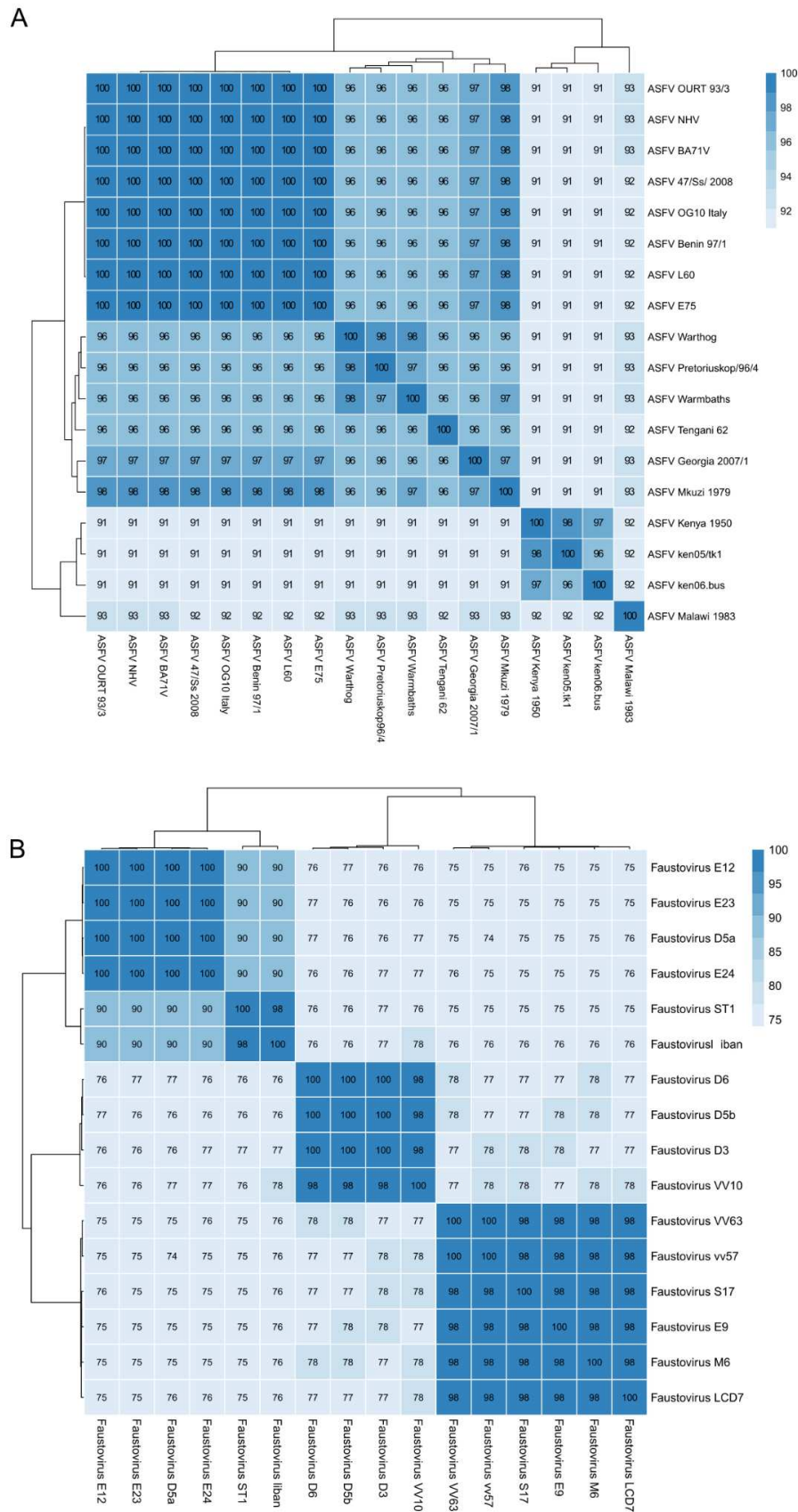
As análises de ANI foram realizadas individualmente para cada grupo, devido à sua limitação metodológica ao comparar grupos com alta divergência entre si. Essa limitação, aliada

à sua resolução, torna-o mais indicado para análises a nível de espécie, sendo considerados de mesma espécie os organismos com genoma acima de 95% de ANI (Goris et al., 2007). Considerando os ASFV, os subgrupos identificados foram os seguintes: 1- ASFV Malawi 1983; 2- ASFV ken05/tk1, Kenya 1950, ken06.Bus; 3- ASFV Georgia 2007/1, Mkuzi 1979, Tengani 62, Warmbaths, Pretoriouskop/96/4 e Warthog; e 4- englobando todos os outros ASFV, formando o terceiro subgrupo (**figura 26A**). Todos esses grupos acomodam genomas acima de 95% de ANI.

Por outro lado, os Faustovírus constituem um grupo ainda mais divergente em termos de ANI do que os ASFV. Os quatro grupos identificados na análise de AAI permanecem consistentes na análise de ANI, com valores superiores a 98% intragrupo (**figura 26B**). No entanto, as comparações intergrupos revelam valores significativamente mais baixos, inferiores a 78%, aproximando-se do limite de confiabilidade da técnica.

Há uma notável divergência genômica entre Kaumoebavirus KLCC10 e Kaumoebavirus Sc, com valores de ANI significativamente baixos, atingindo 77,93%. Em contraste, os Pacmanvirus A23 e Pacmanvirus S19 exibem um nível ligeiramente mais elevado de ANI, alcançando 85,64%. Essa diferença sugere uma maior similaridade genômica entre os dois pacmanvírus em comparação com os dois kaumoebavírus entretanto, ambas comparações abaixo de 95% de ANI.

**Figura 26** – Heatmap de valores de identidade média de nucleotídeos (ANI) par-a-par entre (A) 18 genomas de *Asfvirus* e (B) 16 genomas de *Faustovirus*. ASFV – *African Swine Fever Virus*.

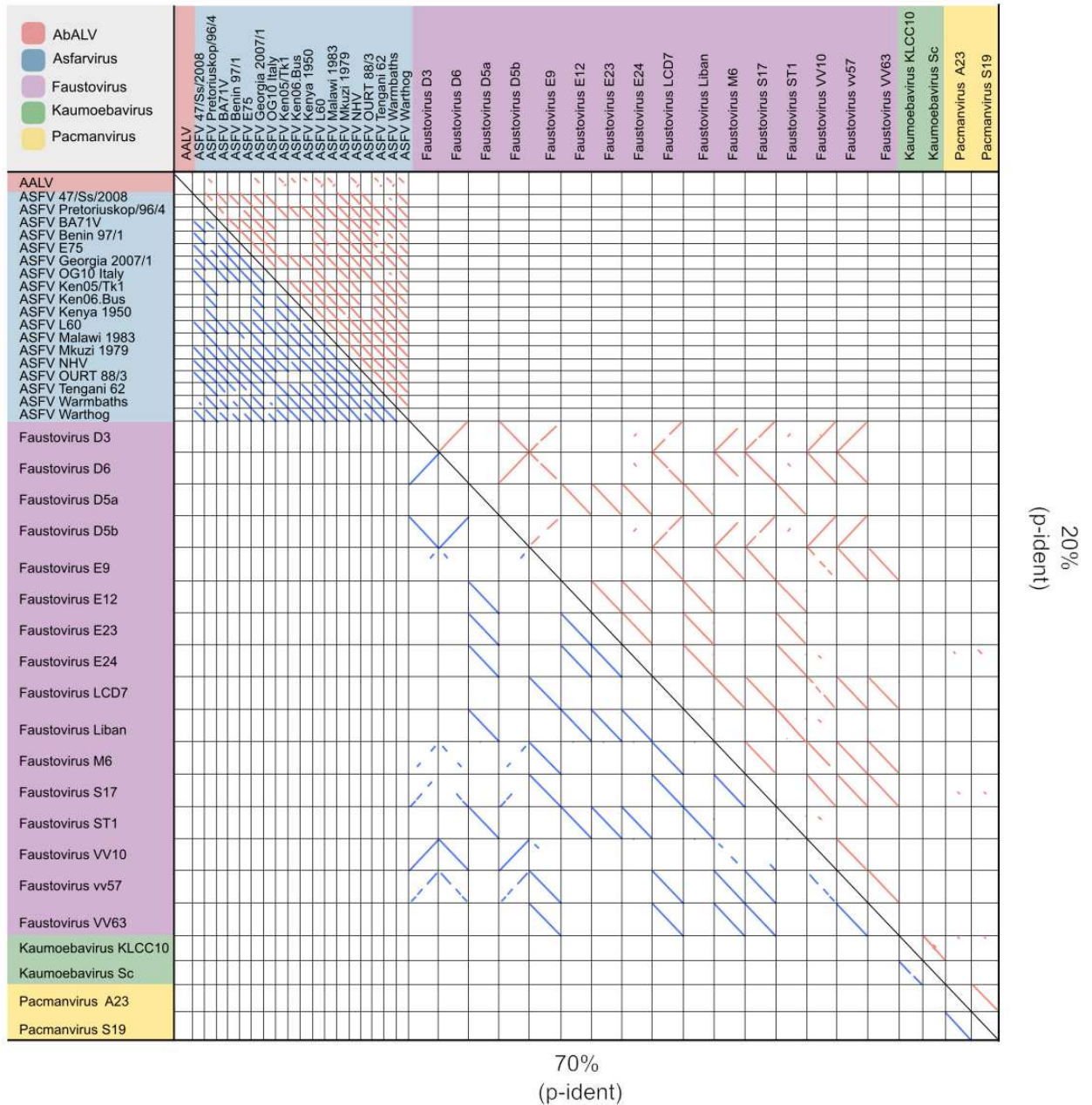


#### 5.4 ANÁLISE DE SINTENIA REVELA VARIAÇÕES NA CONSERVAÇÃO DA ORDEM GÊNICA EM VÍRUS DE EXTENDED *Asfarviridae*

Análises de sintenia em *dot-plot* oferecem uma visão abrangente da organização genômica. Foi empregado BLASTp em dois limiares de identidade (20 e 70%) em conjunto com MCSanX, ferramenta que detecta sintenia. Os resultados revelaram que a conservação da ordem gênica é particular para cada grupo de vírus, mesmo quando se reduz a similaridade entre as proteínas. Ao comparar intergrupos, observou-se que apenas AbALV e ASFV apresentaram colinearidade gênica por meio de blocos gênicos segmentados no genoma (**figura 27**), especialmente quando a similaridade de proteína (*p-ident*) foi diminuída.

Dentro dos grupos, os ASFV demonstraram alta conservação da ordem gênica, exceto pelo grupo formado por ASFV BA71V, Benin 97/1 e E75, que não mostrou colinearidade com o grupo formado por ASFV ken05/tk1, Kenya 1950, ken06.Bus. Por outro lado, os Faustovírus exibiram variações na manutenção da ordem gênica, indo desde a completa ausência até a conservação total, dependendo dos membros específicos do grupo. Ainda em faustovírus, há o fenômeno da inversão das sequências genômicas, como visto nas linhas diagonais indo da parte inferior esquerda para a superior direita. Caixas brancas representam ausência de similaridade entre o par analisado.

**Figura 27** – Análise de sintenia de 39 membros da família *extended Asfarviridae* (Linhas azuis: identidade mínima de 70%, Linhas vermelhas: identidade mínima de 20%). Cada bloco correspondente representa a comparação de colinearidade de dois vírus. Se não houver sequência de aminoácidos colinear entre dois vírus, o bloco estará em branco. AbALV – *Abalone asfar-like virus*; ASFV – *African Swine Fever Virus*.





## 6. DISCUSSÃO

A denominação "*extended Asfarviridae*" surgiu como uma classificação temporária para abarcar uma variedade de vírus gigantes, incluindo AbALV, faustovírus, kaumoebavírus e pacmanvírus, além de possíveis outros vírus a serem descobertos que apresentem algum nível de similaridade filogenética com os ASFV, os membros originais desse clado. A inclusão desses novos membros resulta em uma ampla diversidade biológica dentro do *extended Asfarviridae*. Estes vírus têm a capacidade de infectar uma variedade de hospedeiros, tanto eucariotos unicelulares quanto multicelulares, em ambientes marinhos e terrestres. Suas implicações abrangem áreas como saúde animal, evolução viral, ecologia viral e microbiana.

Compreender adequadamente os ASFV e AbALV, especialmente no contexto da peste suína africana e amiotrofia de abalones, é vital devido às enormes perdas econômicas que essas doenças causam à indústria alimentícia e agropecuária, decorrentes da morte de porcos domésticos e abalones (Matsuyama et al., 2021; Penrith & Vosloo, 2009). A taxonomia de vírus desempenha um papel crucial nesse entendimento, pois permite que as características clínicas, biológicas e evolutivas de um vírus sejam organizadas em uma estrutura que acomoda e conecta todos os vírus, fornecendo assim um valor prático inestimável. Além disso, os vírus de eucariotos unicelulares presentes nesse grupo, como os faustovírus, kaumoebavírus e pacmanvírus, desempenham um papel importante no entendimento da evolução viral e da ecologia dos ambientes aquáticos e terrestres.

Uma taxonomia precisa é fundamental para entender as relações evolutivas entre esses vírus e seus hospedeiros, bem como para explorar seu potencial impacto sobre os ecossistemas. A taxonomia incorreta pode levar a interpretações errôneas dos dados genômicos quando utilizados em estudos mais abrangentes ou até mesmo estudo mais focalizados (Almeida et al., 2023; Bortolus, 2008), prejudicando a compreensão da biologia viral e comprometendo as estratégias de controle viral. Análises, estudos e políticas baseados em classificações inadequadas podem resultar em ações ineficazes ou mesmo prejudiciais para a saúde pública e o meio ambiente e interpretações evolutivas equivocadas (Vogel Ely et al., 2017).

A análise dos genomas dos vírus pertencentes ao grupo *extended Asfarviridae* revelou padrões distintos de diversidade genômica e evolutiva. Os atributos dos genomas foram analisados, incluindo o comprimento do genoma, conteúdo G+C e quantidade de proteínas preditas. Esses padrões genômicos fornecem informações rápidas e valiosas para a classificação e compreensão genômica viral. O conteúdo de G+C varia entre diferentes organismos e se correlaciona com características da história evolutiva do vírus (Romiguier et al., 2010). Esse parâmetro é frequentemente usado em descrições taxonômicas de espécies e gênero, variando

não mais que 1% entre espécies (Meier-Kolthoff et al., 2014). Além disso, o par de base G+C aumenta o acúmulo de energia no genoma do organismo, tornando o DNA mais suscetível a mutações. Hipotetiza-se que, com o tempo, as espécies tendem a diminuir o conteúdo G+C, tornando o DNA mais estável (Chen & Skylaris, 2021). O estudo por Witt et al., (2024) revela que o conteúdo G+C de vírus em famílias de *Nucleocyotoviricota* é majoritariamente discrepante, com exceção de uma maior uniformidade encontrada entre os membros de *Marseilleviridae*. Esse trabalho demonstra na prática que o conteúdo G+C provavelmente não é um bom indicador taxonômico de família em *Nucleocyotoviricota*. Nesse contexto, os faustovírus se aproximam de *Asfvirus*, enquanto os outros grupos divergem entre si. Ainda em faustovírus, a separação dos genomas em três faixas principais de conteúdo G+C reforça a ideia da separação de grupos dentro de faustovírus.

Já a quantidade de proteínas codificadas também pode estar relacionada com a taxonomia em virologia. O número de proteínas codificadas não costuma variar muito dentro do mesmo grupo viral (Rangel-Pineros et al., 2023). Este padrão é observado nos membros de faustovírus, pacmanvírus e kaumoebavírus, onde o número de proteínas codificadas se mantém consistente. Ainda, há pouca divergência entre eles, portanto, o número de proteínas codificadas é um fator que une esses três grupos. Por outro lado, ASFV diverge de todos os grupos ao codificar a menor quantidade de proteínas.

Embora a quantidade de proteínas esteja correlacionada com o tamanho do genoma, especialmente em vírus, a densidade de codificação pode variar. Nos faustovírus, kaumoebavírus e pacmanvírus, essa variação está relacionada principalmente à densidade gênica, com um número semelhante de proteínas sendo codificadas por genomas de tamanhos distintos. Em termos de tamanho do genoma, a diferença significativa entre ASFV, AbALV, kaumoebavírus+pacmanvírus e faustovírus evidencia suas disparidades. O faustovírus, com o maior genoma médio, é 2,5 vezes maior que o genoma médio de ASFV, que possui o menor tamanho genômico do grupo. Genomas virais de membros de mesma família taxonômica tendem a apresentar consistência em seu comprimento (Yuan et al., 2022).

O pangenoma, representado pela totalidade dos genes de uma espécie, é uma fonte crucial de conhecimento ainda não totalmente explorada. A conservação de funções biológicas dos ortólogos indica seus papéis essenciais na sobrevivência e adaptação dos organismos (Koonin, 2005; Padró et al., 2022; Willger et al., 2012). Compreender o pangenoma é essencial para desvendar a diversidade genética intraespecífica, impulsionando a compreensão da evolução, adaptação e interações ecológicas das espécies. Uma análise anterior do pangenoma

de *extended Asfarviridae* por Karki et al. (2021) focou principalmente a investigação em *Metagenomic Assembled Genomes* (MAG) aquáticos, e sem estratificar o pangenoma em seus diversos grupos ou calcular a natureza aberta ou fechada do pangenoma. No presente estudo, uma abordagem mais detalhada foi adotada, agrupando 39 genomas de espécies isoladas de *extended Asfarviridae*, examinando minuciosamente cada grupo pangenômico e investigando suas características distintivas, tanto em termos de composição quanto de funcionalidade das proteínas.

A literatura científica não oferecia uma definição precisa acerca da natureza, aberta ou fechada, do pangenoma de *extended Asfarviridae*. Este estudo abordou essa questão mediante a aplicação da fórmula de Heaps, proposta por Tettelin et al. (2005), e pela elaboração de uma curva de acumulação pangenômica, com o intuito de avaliar este aspecto. Os resultados indicam que o pangenoma do *extended Asfarviridae* é considerado aberto. Tal categorização implica que a descoberta e incorporação de novos genomas tendem a acrescentar novos genes ao pangenoma, insinuando, assim, que a diversidade de *extended Asfarviridae* ainda não é completamente conhecida. Em contrapartida a muitos grupos virais, cujos pangenomas são fechados devido ao número limitado de genes, os NCLDV podem apresentar genomas desprovidos dessa limitação, resultando em pangenomas abertos, como o observado neste estudo. Exemplos de outros grupos virais com pangenomas abertos incluem os Pandoravirus (Aherfi et al., 2018), a família *Herpesviridae* (Sharma et al., 2016) e os próprios ASFV a nível de espécie (L. Wang et al., 2020).

O pangenoma aberto de *extended Asfarviridae* indica sua expansão contínua através de diferentes estratégias de aquisição gênica. Essa natureza aberta é consistente com a hipótese de que espécies que habitam uma ampla gama de ambientes tendem a ter um pangenoma aberto (Medini et al., 2005; Tettelin et al., 2005), estando em linha com os vírus de *extended Asfarviridae*, os quais infectam organismos unicelulares e multicelulares em ambientes aquáticos e terrestres. Além disso, sugere-se que o estado aberto do pangenoma pode estar ligado à instabilidade genômica (Rajput et al., 2023), associação essa que está em conformidade com a alta proporção de G+C nas espécies de *extended Asfarviridae*, o qual está relacionado com aumento da mutação (Chen & Skylaris, 2021).

Na análise da curva pangenômica, o grupo de genes *core* revela-se conservado e com pouca variação à medida que novos genomas são incorporados, estabilizando rapidamente após uma redução abrupta, evidenciando a alta dissimilaridade entre esses genomas. Além disso, trata-se de um genoma *core* diminuto nesse clado, composto por apenas 37 genes, em comparação com o tamanho do pangenoma, que totaliza 2.483 clusters de ortólogos. Tal

proporção é inferior à observada nos genomas *core* de outros vírus gigantes, como os Pandoravirus, com 425 genes *core* correspondendo de 15 a 30% do genoma total das espécies (Aherfi et al., 2018), *Mimiviridae*, com 267 genes *core* (Assis et al., 2017) correspondendo a 22 a 28% do genoma de espécies dessa família e *Marseilleviridae*, com 202 genes *core* (Dornas et al., 2016), representando cerca de 44% do genoma total. A questão da classificação taxonômica dos vírus pertencentes ao clado dos pithovirus, que engloba cedratvírus e orpheovírus, também tem sido objeto de debate na literatura científica. Em um estudo conduzido por Queiroz et al. (2023), por meio de uma análise genômica comparativa abrangente, foi estabelecido um conjunto de 52 genes *core* essenciais para esse grupo, número mais próximo de *extended Asfarviridae*. Essa investigação sugere que a unidade taxonômica desses vírus como uma família coesa pode ser questionável.

Destaca-se a observação que o genoma *core* de *extended Asfarviridae* se assemelha ao dos NCLDV's em número e composição, os quais apresentam 47 genes no genoma ancestral e atualmente 40 genes (Yutin et al., 2009). Essa observação sugere que a comparação do conteúdo gênico realizada aqui tende a alcançar o nível basal do genoma *core* desse grande grupo, evidenciando a comparação de elementos distintos uns dos outros. Se os membros de *extended Asfarviridae* fossem mais geneticamente próximos entre si, é razoável supor que o genoma *core* seria mais amplo. Por exemplificação, o genoma *core* do gênero *Asfivirus* é composto por 86 genes (L. Wang et al., 2020), 2,3 vezes maior que o genoma *core* de *extended Asfarviridae*, e correspondente a 53% do genoma de vírus desse gênero. Essa observação sugere que o tamanho do genoma *core* pode servir como uma métrica útil para avaliar a proximidade genética entre os vírus de determinado grupo.

Para cada linhagem analisada em *extended Asfarviridae*, os genes *core* representam aproximadamente 7~8% do total nos faustovírus, kaumoebavírus e pacmanvírus. Entretanto, essa proporção varia em AbALV, onde os genes *core* compõem 10% do total, e nos ASFV, variando de 20 a 25% do genoma. Essa disparidade reflete não apenas a diferença de tamanho entre os genomas, com os ASFV apresentando genomas consideravelmente menores, mas também indica a magnitude do genoma acessório. Os membros do *extended Asfarviridae* exibem um genoma acessório robusto, sugerindo um amplo arsenal de genes adaptativos que potencialmente conferem vantagens na interação com a diversidade de ambientes e hospedeiros desse grupo além de servirem de berço para adaptação em novos ambientes (Álvarez et al., 2020; Croll & McDonald, 2012). Isso fica claro ao comparar com o tamanho do genoma acessório menor das outras famílias de NCLDV aqui citados, possuindo um espectro tanto de hospedeiros como de ambiente mais estrito.

A distribuição do tamanho das proteínas por grupo pangenômico revela uma notável dissimilaridade. As proteínas do grupo *core* demonstram uma tendência a serem maiores em comparação com os outros grupos, sugerindo sua importância funcional e evolutiva. Esta característica pode indicar que as proteínas do grupo *core* são potencialmente mais antigas, uma vez que proteínas mais longas tendem a ter uma origem evolutiva mais remota (Toll-Riera & Albà, 2013). Além disso, dado que a seleção natural tende a suprimir mudanças em genes com transcrições mais longas e promover mudanças em genes com transcrições menores (Lopes et al., 2021), isso ressalta a relevância das proteínas deste grupo. Em contrapartida, as proteínas do grupo *cloud* tendem a ser menores, o que evidencia sua natureza acessória e sugere uma idade mais recente, além de uma maior propensão a acumular mutações.

Uma parcela significativa das proteínas presentes no pangenoma de *extended Asfarviridae* carece de função conhecida, uma desvantagem recorrente nos genomas dos NCLDV. Desde sua identificação inicial, uma diversidade impressionante de genes, até então desconhecidos na natureza, tem sido descoberta, com relatos de NCLDV apresentando mais de 80% do genoma composto por proteínas não caracterizadas (Rigou et al., 2022). O grupo com a maior proporção de proteínas cuja função foi identificada é o grupo *core*, devido à sua conservação entre diversas espécies e à presença de genes essenciais já documentados, fatores que possivelmente contribuem para a identificação da maioria das proteínas nesse grupo. O próprio processo metodológico adotado para atribuição de funções enfatiza a dificuldade intrínseca em anotar funcionalmente essas proteínas, exigindo uma abordagem em três etapas distintas: processamento em larga escala, com o eggNOG-mapper, curadoria manual, com o BLASTp, e transição da análise de sequências para análise de estruturas proteicas, com AlphaFold e FoldSeek.

Sabe-se que diferentes sequências de aminoácidos podem adotar uma mesma estrutura tridimensional por meio de evolução convergente; inclusive, sequências com similaridade abaixo de 8% podem apresentar estruturas semelhantes (Rost, 1997). Na biologia estrutural, a estrutura proteica é considerada um indicativo de função. Essa abordagem mostrou-se eficaz ao identificar a possível função de duas das quatro proteínas hipotéticas presentes no genoma *core*: uma subunidade da RNA polimerase II e um domínio quinase. As estruturas preditas das proteínas hipotéticas apresentam um *fold* similar a partes dessas proteínas homólogas. O fato de a possível subunidade de RNA polimerase II aqui predita mostrar-se similar à estrutura de uma proteína de um poxvírus, um vírus proximoamente relacionado a *extended Asfarviridae*, valida ainda mais o modelo. Isso nos leva a considerar como a história evolutiva desses vírus gerou estruturas proteicas semelhantes às encontradas em maquinarias de replicação canônicas

na biosfera a partir de sequências primárias diferentes. Além disso, suscita reflexões sobre a importância dessas estruturas, uma vez que continuam a ser formadas mesmo com sequências aminoacídicas distintas. Esse método oferece uma perspectiva promissora para identificar funções de proteínas desconhecidas em vírus gigantes e outras espécies dado a limitação encontrada com busca de homologia por sequência.

A maioria das proteínas identificadas no pangenoma de *extended Asfarviridae*, especialmente no genoma *core*, desempenha funções relacionadas aos processos do DNA, como replicação, recombinação ou reparo, exemplificadas pela presença de genes como a *DNA polimerase da família B* e a *recombinase YqaJ*. Ademais, observa-se a presença de genes envolvidos no processamento de transcrito, como os que codificam enzimas responsáveis pela remoção do cap5' de mRNAs, e de proteínas metabólicas, como a metiltransferase do tipo Ftsj e a protease Ulp1, além de componentes relacionados à transcrição e processamento de RNA, como a RNA polimerase e os fatores iniciadores da transcrição IIB e S-II. Tais genes, comumente denominados *housekeeping*, são essenciais para a manutenção biológica básica do genoma desses vírus, especialmente em se tratando de vírus de DNA e, muitos deles, incluindo todos os genes citados, estão presentes no conjunto de genes *core* dos NCLDV (Yutin et al., 2009).

É importante ressaltar que o gene da *Major Capsid Protein* (MCP) não foi contabilizado no grupo *core*, embora seja essencial para a estrutura de todos os vírus do *extended Asfarviridae*, pois no caso do faustovírus, este gene encontra-se fragmentado ao longo do genoma (Cherif Louazani et al., 2018). Essa ausência não é incomum em genomas *core*; em diversos estudos, ele também não é detectado em Pandoravirus e em alguns pithovirus divergentes (Rigou et al., 2022). Tal limitação da técnica de pangenoma evidencia que, em situações em que a estrutura de um determinado gene foge do padrão convencional, sua presença pode não ser identificada.

Os demais grupos pangenômicos apresentaram uma anotação funcional limitada, devido ao grande número de proteínas desconhecidas no pangenoma. Entretanto, entre as funções identificadas, observa-se uma predominância de processos relacionados à replicação, recombinação e reparo de DNA, transporte e metabolismo de aminoácidos, especialmente no grupo do *soft core*, e transporte e metabolismo de nucleotídeos, modificação pós-traducional, reciclagem de proteínas e chaperonas. Devido à elevada quantidade de genes *singletons* no grupo *cloud*, era esperado um baixo nível de anotação funcional, praticamente insignificante.

A análise filogenética dos genes *core*, mais consistente com evolução viral do que filogenia de gene único (Yu et al., 2021), juntamente com a avaliação de identidade de

aminoácidos (AAI), identidade de nucleotídeos (ANI) e sintenia compõem um conjunto de fatores cruciais para compreender a proximidade genética dentro do grupo *extended Asfarviridae*. A filogenia, baseada em genes essenciais ao genoma viral, é o principal critério taxonômico, fornecendo uma representação precisa das relações evolutivas entre os vírus (Gorbalenya & Lauber, 2017). AAI e a ANI, métricas comumente usadas, complementam essa análise, permitindo comparações genômicas mais detalhadas e identificando relações genéticas entre as espécies (Rodriguez-R & Konstantinidis, 2016). A análise de sintenia, que examina a conservação da organização e ordem genômica, também é empregada, oferecendo percepções adicionais sobre a evolução e a diversidade genômica do grupo (Y. Wang et al., 2012). A integração dessas métricas permite uma avaliação abrangente da formação de táxons, revelando padrões evolutivos e relacionamentos genéticos mais complexos.

A análise dos 39 genomas de *extended Asfarviridae* revela uma notável diversidade genética com maior intensidade intergrupos que intragrupo. A filogenia dos genes *core* destaca as distâncias evolutivas entre os grupos virais, com uma considerável separação dos faustovírus, kaumobavírus, pacmanvírus e AbALV em relação aos ASFV. A utilização do AAI, recomendada pelo ICTV para delimitações taxonômicas em nível de família, revelou similaridade global de 49%, apesar de alta variabilidade, indicando uma identidade comum. No entanto, a presença de subgrupos com valores extremos de AAI em contraste com regiões de baixa identidade sugere uma composição aminoacídica não uniforme nesse clado. Os resultados da análise de sintenia também demonstram essa heterogeneidade, com manutenção da colinearidade intragrupo mais comum do que intergrupos, mesmo em valores de identidade (*p-ident*) extremos, destacando uma história evolutiva distinta e bem delimitada para cada subgrupo.

Os vírus da peste suína africana (ASFV), os únicos membros oficialmente reconhecidos da família *Asfarviridae*, exibem uma proximidade genética excepcional entre si, formando um grupo uniforme e coeso com uma similaridade extremamente alta, especialmente quando avaliados através da AAI, com valores superiores a 95%. Esse valor pode ser utilizado como limiar na classificação de novos vírus no gênero *Asfivirus*. Além disso, os resultados de análises de sintenia revelam uma notável colinearidade genômica entre os diferentes membros deste grupo. Essa coesão genética se manifesta na formação de politomias na árvore filogenética, sugerindo uma equidistância genética predominante entre os diversos membros. No entanto, observa-se uma ligeira separação na linhagem composta pelos isolados ASFV ken05/tk1, Kenya 1950 e ken06.Bus, os quais também compartilham uma baixa ordem genômica com os demais membros do grupo. Adicionalmente, a linhagem ASFV Malawi 1983 se destaca

consistentemente nas análises como levemente divergente em relação aos outros membros. As análises de ANI corroboram esses resultados, sugerindo a possível subdivisão dos ASFV em três espécies distintas, caso se adote um critério de similaridade genética superior a 95% de ANI (Rodriguez-R & Konstantinidis, 2016). Sob essa classificação, somados aos resultados da filogenia e colinearidade, ASFV ken05/tk1, Kenya 1950 e ken06.Bus seriam designados como isolados de uma espécie, enquanto Malawi 1983 constituiria um isolado de outra. Por fim, ASFV Georgia 2007/1, Mkuzi 1979, Tengani 62, Warmbaths, Pretoriouskop/96/4, Warthog, E75, L60, Benin 97/1, OG10 Italy, 47/Ss/2008, BA71V, NIHV e OURT 93/3 seriam considerados como isolados de uma terceira espécie do gênero *Asfivirus*.

Os faustovírus não exibem uma uniformidade tão marcante em termos de análise filogenética como os ASFV, evidenciando uma clara separação evolutiva entre eles. A análise de AAI revela a formação de quatro grupos altamente homogêneos entre si, um padrão que também é observado na análise de ANI. Esses achados estão em concordância com as quatro linhagens de faustovírus previamente definidas por Benamar et al., (2016).

Um limiar acima de 65% de AAI é frequentemente observado para membros do mesmo gênero (Konstantinidis et al., 2017). Quando comparados intergrupos dentro das quatro linhagens de faustovírus, esse valor varia entre 50% e 60%, indicando estar em um limite crítico quanto à delimitação de gênero. A organização genômica também é influenciada pela linhagem de faustovírus, mantendo-se mais preservada quando os genomas comparados são da mesma linhagem.

A combinação dos dados de conteúdo G+C, AAI, ANI, filogenia dos genes *core* e sintenia sugere que os faustovírus se dividem em quatro espécies virais distintas: Espécie 1 (Faustovírus E12, D5a, E23, E24), Espécie 2 (Faustovírus ST1 e Liban), Espécie 3 (E9, M6, S17, LCD7, vv57, VV63) e Espécie 4 (D6, D5b, D3, VV10). Entretanto, ao considerar um limiar entre 95% e 90% de ANI para subespécie, Faustovírus Espécie 1 pode ser considerado uma subespécie de Faustovírus Espécie 2. No entanto, muitos taxonomistas argumentam pela extinção desse táxon devido à descontinuidade genética observada entre esses organismos gerando limites mal estabelecidos e por tornar-se irrelevantes e de valor limitado para estudos que utilizam de *frameworks* taxonômicos (Venter et al., 2022). Além disso, há a possibilidade de que diferentes gêneros virais estejam presentes, conforme sugerido pela análise de AAI, o que requer investigação adicional a fim de encontrar características fenotípicas que consigam diferenciá-los

Os kaumobavírus e pacmanvírus demonstram consistentemente uma maior uniformidade dentro de seus próprios grupos do que entre grupos, indicando uma proximidade



genômica, entretanto, não equidistante. As evidências sugerem que, apesar da colinearidade genômica observada ao longo da maior parte de seus genomas, os dois kaumobavírus pertencem a espécies diferentes, conforme indicado pela análise de ANI, e estão no limite para serem classificados no mesmo gênero com base na AAI. Por outro lado, os pacmanvírus mostram uma maior proximidade entre si em comparação com os kaumobavírus, com valores de AAI e ANI que sugerem que são do mesmo gênero, mas de espécies diferentes.

O AbALV emerge como o membro mais próximo dos ASFV em todas as análises realizadas. É o único genoma que exhibe colinearidade com os ASFV, embora essa colinearidade exija uma diminuição na identidade (*p-ident*) para se manifestar. Além disso, esses dois grupos são os únicos do clado a infectar eucariotos multicelulares, o que reforça sua proximidade evolutiva.

## 7. CONCLUSÃO

O grupo de vírus gigantes “*extended Asfarviridae*” apresenta uma composição de vírus geneticamente relacionados até certo ponto. O pangenoma revela um compartilhamento, embora reduzido, de genes ortólogos entre diversos vírus deste clado, além de um conjunto restrito de genes *core* compartilhados por todos os membros do grupo. Similaridades também são observadas na forma e estruturação do capsídeo para a maioria dos membros, bem como no hospedeiro para dois dos cinco vírus analisados. Contudo, as semelhanças limitam-se a esses aspectos.

A análise comparativa detalhada evidencia uma complexidade genômica notável, caracterizada por diferenças significativas em seus genomas, tanto em termos de composição de aminoácidos quanto de nucleotídeos. Observa-se um conjunto reduzido de genes *core* em comparação com outras famílias virais, divergências filogenéticas acentuadas e ausência de manutenção da sintenia. Além disso, existem variações no tamanho do genoma, no conteúdo de G+C, no número de proteínas codificadas, bem como nas características fenotípicas, incluindo o hospedeiro celular e a patologia associada.

Esses achados sugerem que esses vírus agregados provavelmente não se enquadram em uma única família taxonômica, principalmente quando se comparados ao único membro oficial, os vírus do gênero *Asfivirus*. Faustovírus, kaumoebavírus, pacmanvírus e *Abalone asfar-like virus* representam um agrupamento de espécies próximas, mas geneticamente diversas, que provavelmente formarão suas próprias famílias. Nesse aspecto, também foi sugerida a divisão dos 16 faustovírus em 4 espécies distintas, conseguimos sugerir a divisão dos 18 *Asfivirus* em três espécies distintas e sugerir que as duplas de pacmanvírus e kaumoebavírus também formariam quatro espécies distintas.

Há uma baixa resolução taxonômica que pode ser atribuída, em parte, à falta de isolamentos virais representativos o suficiente para capturar toda a diversidade genética desse grupo. Percebe-se, pela natureza aberta desse pangenoma, que a diversidade nesse grupo ainda não se esgotou. Portanto, a obtenção de mais sequências virais e a realização de estudos adicionais são cruciais para aprimorar nossa compreensão da diversidade genômica e evolutiva desses vírus, bem como sua relação com os hospedeiros. Possivelmente, a descoberta de novos faustovírus, kaumoebavírus, pacmanvírus e *Abalone asfar-like virus* suscite a formação de novas famílias virais, a fim de acomodar esses membros.

Essa análise ressalta a importância da amostragem taxonômica adequada e a necessidade contínua de investigação para elucidar de forma mais precisa a posição filogenética desses vírus. Assim, podemos antecipar que com a descoberta de genomas virais gigantes

adicionais, a posição filogenética desses vírus será melhor resolvida, contribuindo para um entendimento mais abrangente da diversidade viral e sua posição nos ecossistemas.

## 8. PERSPECTIVAS

Este estudo oferece uma visão abrangente das análises genômicas e funcionais, contribuindo significativamente para o entendimento dos vírus gigantes, em especial de *extended Asfarviridae*, um campo ainda em expansão com vasto potencial de pesquisa. Ao estabelecer e descrever o conjunto de genes *core* desses vírus, adicionamos uma camada de compreensão crucial para a evolução desse grupo viral.

Para aprimorar ainda mais nosso conhecimento, abordagens adicionais podem ser exploradas, como a investigação das regiões não codificantes e a construção de um pangenoma de sequências nucleotídicas dessas regiões a fim de investigar o compartilhamento de sequências regulatórias e *motifs*. Além disso, a incorporação de dados genômicos de vírus recém-descobertos têm o potencial de fornecer informações valiosas para resolver questões pendentes sobre a classificação taxonômica desses vírus.

Paralelamente às análises genômicas, é fundamental realizar estudos funcionais nas proteínas desses vírus para compreender melhor suas características biológicas. Essa abordagem pode enriquecer nosso entendimento da diversidade global de proteínas e até mesmo gerar proteínas com aplicações biotecnológicas. Além disso, a identificação de funções proteicas pode lançar luz sobre os fatores de virulência dos *Asfvirus* e *Abalone asfar-like virus*, que são patogênicos. A utilização da biologia estrutural para definir a função das proteínas mostra-se como uma ferramenta complementar valiosa, especialmente em estudos de proteínas desconhecidas de vírus gigantes. Explorar a diversidade funcional das proteínas do grupo *cloud* e investigar a origem de novos genes também são áreas promissoras para futuras pesquisas em virologia e nascimento de novos genes.

## REFERÊNCIAS

*A global battle against African swine fever—Global Times.* (2019).

<https://www.globaltimes.cn/page/201909/1164818.shtml>

A. Marouf Et Al., M. (2021). Acanthamoeba genotypes and Vermamoeba vermiformis in drinking water treatment facilities. *Egyptian Journal of Aquatic Biology and Fisheries*, 25(6), 327–341. <https://doi.org/10.21608/ejabf.2021.212783>

ABEDKHOJASTEH, H., NIYYATI, M., RAHIMI, F., HEIDARI, M., FARNIA, S., & REZAEIAN, M. (2013). First Report of Hartmannella keratitis in a Cosmetic Soft Contact Lens Wearer in Iran. *Iranian Journal of Parasitology*, 8(3), 481–485.

Adams, M. J., Hendrickson, R. C., Dempsey, D. M., & Lefkowitz, E. J. (2015). Tracking the changes in virus taxonomy. *Archives of Virology*, 160(5), 1375–1383.

<https://doi.org/10.1007/s00705-015-2376-4>

Aherfi, S., Andreani, J., Baptiste, E., Oumessoum, A., Dornas, F. P., Andrade, A. C. dos S. P., Chabriere, E., Abrahao, J., Levasseur, A., Raoult, D., La Scola, B., & Colson, P.

(2018). A Large Open Pangenome and a Small Core Genome for Giant Pandoraviruses. *Frontiers in Microbiology*, 9, 1486.

<https://doi.org/10.3389/fmicb.2018.01486>

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

<https://doi.org/10.1016/j.virol.2014.07.014>

Alempic, J.-M., Lartigue, A., Goncharov, A. E., Grosse, G., Strauss, J., Tikhonov, A. N., Fedorov, A. N., Poirot, O., Legendre, M., Santini, S., Abergel, C., & Claverie, J.-M. (2023). An Update on Eukaryotic Viruses Revived from Ancient Permafrost. *Viruses*, 15(2), Artigo 2. <https://doi.org/10.3390/v15020564>

Almeida, R. F. de, Pellegrini, M. O. O., Morais, I. L. de, Simão-Bianchini, R.,

Rattanakrajang, P., Cheek, M., & Simões, A. R. G. (2023). Barking up the wrong tree:

- The dangers of taxonomic misidentification in molecular phylogenetic studies. *Plant Ecology and Evolution*, 156(2), Artículo 2. <https://doi.org/10.5091/plecevo.101135>
- Alonso, C., Borca, M., Dixon, L., Revilla, Y., Rodriguez, F., Escribano, J. M., & ICTV Report Consortium. (2018). ICTV Virus Taxonomy Profile: Asfarviridae. *Journal of General Virology*, 99(5), 613–614. <https://doi.org/10.1099/jgv.0.001049>
- Álvarez, V. E., Quiroga, M. P., Galán, A. V., Vilacoba, E., Quiroga, C., Ramírez, M. S., & Centrón, D. (2020). Crucial Role of the Accessory Genome in the Evolutionary Trajectory of *Acinetobacter baumannii* Global Clone 1. *Frontiers in Microbiology*, 11, 342. <https://doi.org/10.3389/fmicb.2020.00342>
- Andreani, J., Khalil, J. Y. B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., Colson, P., Klose, T., Rossmann, M. G., Raoult, D., & La Scola, B. (2017). Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads between Asfarviridae and Faustoviruses. *Journal of Virology*, 91(14), e00212-17. <https://doi.org/10.1128/JVI.00212-17>
- Andrés, G., Charro, D., Matamoros, T., Dillard, R. S., & Abrescia, N. G. A. (2020). The cryo-EM structure of African swine fever virus unravels a unique architecture comprising two icosahedral protein capsids and two lipoprotein membranes. *The Journal of Biological Chemistry*, 295(1), 1–12. <https://doi.org/10.1074/jbc.AC119.011196>
- Andrés, G., Simón-Mateo, C., & Viñuela, E. (1997). Assembly of African swine fever virus: Role of polyprotein pp220. *Journal of Virology*, 71(3), 2331–2341. <https://doi.org/10.1128/JVI.71.3.2331-2341.1997>
- Arab-Mazar, Z., Niyiyati, M., Lasjerdi, Z., Spotin, A., Alavi Darzam, I., & Gachkar, L. (2019). Isolation, identification, and phylogenetic analysis of potentially pathogenic free-living amoebae isolated from nasal and oral mucosa of HIV/AIDS patients in Iran. *Parasitology Research*, 118(10), 3061–3066. <https://doi.org/10.1007/s00436-019-06448-x>

- Assis, F. L., Franco-Luiz, A. P. M., Dos Santos, R. N., Campos, F. S., Dornas, F. P., Borato, P. V. M., Franco, A. C., Abrahao, J. S., Colson, P., & Scola, B. L. (2017). Genome Characterization of the First Mimiviruses of Lineage C Isolated in Brazil. *Frontiers in Microbiology*, 8, 2562. <https://doi.org/10.3389/fmicb.2017.02562>
- Aylward, F. O., Moniruzzaman, M., Ha, A. D., & Koonin, E. V. (2021). *A Phylogenomic Framework for Charting the Diversity and Evolution of Giant Viruses* (p. 2021.05.05.442809). bioRxiv. <https://doi.org/10.1101/2021.05.05.442809>
- Bäckström, D., Yutin, N., Jørgensen, S. L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., Spang, A., Wolf, Y. I., Koonin, E. V., & Ettema, T. J. G. (2019). Virus Genomes from Deep Sea Sediments Expand the Ocean Megavirome and Support Independent Origins of Viral Gigantism. *mBio*, 10(2), e02497-18. <https://doi.org/10.1128/mBio.02497-18>
- Bajrai, L. H., Benamar, S., Azhar, E. I., Robert, C., Levasseur, A., Raoult, D., & La Scola, B. (2016). Kaumoebavirus, a New Virus That Clusters with Faustoviruses and Asfarviridae. *Viruses*, 8(11), Artigo 11. <https://doi.org/10.3390/v8110278>
- Bandi, V., & Gutwin, C. (2020). *Interactive Exploration of Genomic Conservation* [Application/pdf]. 10 pages, 5090. <https://doi.org/10.20380/GI2020.09>
- Basta, S., Gerber, H., Schaub, A., Summerfield, A., & McCullough, K. C. (2010). Cellular processes essential for African swine fever virus to infect and replicate in primary macrophages. *Veterinary Microbiology*, 140(1–2), 9–17. <https://doi.org/10.1016/j.vetmic.2009.07.015>
- Bellec, L., Grimsley, N., Moreau, H., & Desdevises, Y. (2009). Phylogenetic analysis of new Prasinoviruses (Phycodnaviridae) that infect the green unicellular algae *Ostreococcus*, *Bathycoccus* and *Micromonas*. *Environmental Microbiology Reports*, 1(2), 114–123. <https://doi.org/10.1111/j.1758-2229.2009.00015.x>

- Benamar, S., Reteno, D. G. I., Bandaly, V., Labas, N., Raoult, D., & La Scola, B. (2016). Faustoviruses: Comparative Genomics of New Megavirales Family Members. *Frontiers in Microbiology*, 7, 3. <https://doi.org/10.3389/fmicb.2016.00003>
- Berthe, F. (2020). *The global economic impact of ASF*. 2 p. <https://doi.org/10.20506/bull.2020.1.3119>
- Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, 29(12), 2607–2618. <https://doi.org/10.1093/nar/29.12.2607>
- Bortolus, A. (2008). Error Cascades in the Biological Sciences: The Unwanted Consequences of Using Bad Taxonomy in Ecology. *Ambio*, 37, 114–118. [https://doi.org/10.1579/0044-7447\(2008\)37\[114:ECITBS\]2.0.CO;2](https://doi.org/10.1579/0044-7447(2008)37[114:ECITBS]2.0.CO;2)
- Boyer, M., Madoui, M.-A., Gimenez, G., Scola, B. L., & Raoult, D. (2010). Phylogenetic and Phyletic Studies of Informational Genes in Genomes Highlight Existence of a 4th Domain of Life Including Giant Viruses. *PLOS ONE*, 5(12), e15530. <https://doi.org/10.1371/journal.pone.0015530>
- Bradbury, R. S. (2014). Free-Living Amoebae Recovered from Human Stool Samples in Strongyloides Agar Culture. *Journal of Clinical Microbiology*, 52(2), 699–700. <https://doi.org/10.1128/JCM.02738-13>
- Brandes, N., & Linial, M. (2019). Giant Viruses—Big Surprises. *Viruses*, 11(5), 404. <https://doi.org/10.3390/v11050404>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 421. <https://doi.org/10.1186/1471-2105-10-421>



- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, 17(4), 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Chapman, D. A. G., Tcherepanov, V., Upton, C., & Dixon, L. K. (2008). Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. *The Journal of General Virology*, 89(Pt 2), 397–408. <https://doi.org/10.1099/vir.0.83343-0>
- Chen, H., & Skylaris, C.-K. (2021). Analysis of DNA interactions and GC content with energy decomposition in large-scale quantum mechanical calculations. *Physical Chemistry Chemical Physics: PCCP*, 23(14), 8891–8899. <https://doi.org/10.1039/d0cp06630c>
- Cherif Louazani, A., Baptiste, E., Levasseur, A., Colson, P., & La Scola, B. (2018). Faustovirus E12 Transcriptome Analysis Reveals Complex Splicing in Capsid Gene. *Frontiers in Microbiology*, 9. <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02534>
- Chinchar, V. G., Waltzek, T. B., & Subramaniam, K. (2017). Ranaviruses and other members of the family *Iridoviridae*: Their place in the virosphere. *Virology*, 511, 259–271. <https://doi.org/10.1016/j.virol.2017.06.007>
- Claverie, J.-M., & Abergel, C. (2016). Giant viruses: The difficult breaking of multiple epistemological barriers. *Studies in History and Philosophy of Science Part C: Studies*

*in History and Philosophy of Biological and Biomedical Sciences*, 59, 89–99.

<https://doi.org/10.1016/j.shpsc.2016.02.015>

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)*, 25(11), 1422–1423.

<https://doi.org/10.1093/bioinformatics/btp163>

Colson, P., La Scola, B., Levasseur, A., Caetano-Anollés, G., & Raoult, D. (2017).

Mimivirus: Leading the way in the discovery of giant viruses of amoebae. *Nature Reviews Microbiology*, 15(4), 243–254. <https://doi.org/10.1038/nrmicro.2016.197>

Condit, R. C., Moussatche, N., & Traktman, P. (2006). In a nutshell: Structure and assembly of the vaccinia virion. *Advances in Virus Research*, 66, 31–124.

[https://doi.org/10.1016/S0065-3527\(06\)66002-8](https://doi.org/10.1016/S0065-3527(06)66002-8)

Costa, S. S., Guimarães, L. C., Silva, A., Soares, S. C., & Baraúna, R. A. (2020). First Steps in the Analysis of Prokaryotic Pan-Genomes. *Bioinformatics and Biology Insights*, 14, 1177932220938064. <https://doi.org/10.1177/1177932220938064>

Croll, D., & McDonald, B. A. (2012). The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. *PLoS Pathogens*, 8(4), e1002608.

<https://doi.org/10.1371/journal.ppat.1002608>

*Current ICTV Taxonomy Release | ICTV*. ([s.d.]). Recuperado 21 de fevereiro de 2024, de <https://ictv.global/taxonomy>

Cwynar, P., Stojkov, J., & Wlzlak, K. (2019). African Swine Fever Status in Europe.

*Viruses*, 11(4), 310. <https://doi.org/10.3390/v11040310>

Daaboul, G. G., Freedman, D. S., Scherr, S. M., Carter, E., Rosca, A., Bernstein, D., Mire, C. E., Agans, K. N., Hoenen, T., Geisbert, T. W., Ünlü, M. S., & Connor, J. H. (2017).

- Enhanced light microscopy visualization of virus particles from Zika virus to filamentous ebolaviruses. *PLOS ONE*, *12*(6), e0179728.  
<https://doi.org/10.1371/journal.pone.0179728>
- De Paula Lyra, T. M., Saraiva, V. E. V., Hermida Lage, G. R., & Samarcos, M. S. R. (1986). Eradication of African swine fever from Brazil. *Revue Scientifique Et Technique (International Office of Epizootics)*, *5*(3), 771–787.  
<https://doi.org/10.20506/rst.5.3.261>
- Dixon, L. K., Chapman, D. A. G., Netherton, C. L., & Upton, C. (2013). African swine fever virus replication and genomics. *Virus Research*, *173*(1), 3–14.  
<https://doi.org/10.1016/j.virusres.2012.10.020>
- Dornas, F. P., Assis, F. L., Aherfi, S., Arantes, T., Abrahão, J. S., Colson, P., & La Scola, B. (2016). A Brazilian Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae. *Viruses*, *8*(3), 76. <https://doi.org/10.3390/v8030076>
- Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., & Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (New York, N.Y.)*, *359*(6379), eaar4120.  
<https://doi.org/10.1126/science.aar4120>
- Dreher, T. W. (2010). Viral tRNAs and tRNA-like structures. *Wiley Interdisciplinary Reviews. RNA*, *1*(3), 402–414. <https://doi.org/10.1002/wrna.42>
- Eisenstein, M. (2023). Every base everywhere all at once: Pangenomics comes of age. *Nature*, *616*(7957), 618–620. <https://doi.org/10.1038/d41586-023-01300-w>
- Eustace Montgomery, R. (1921). On A Form of Swine Fever Occurring in British East Africa (Kenya Colony). *Journal of Comparative Pathology and Therapeutics*, *34*, 159–191.  
[https://doi.org/10.1016/S0368-1742\(21\)80031-4](https://doi.org/10.1016/S0368-1742(21)80031-4)

- Ezzatpour, S., Mondragon Portocarrero, A. D. C., Cardelle-Cobas, A., Lamas, A., López-Santamarina, A., Miranda, J. M., & Aguilar, H. C. (2023). The Human Gut Virome and Its Relationship with Nontransmissible Chronic Diseases. *Nutrients*, *15*(4), 977. <https://doi.org/10.3390/nu15040977>
- Federici, B. A., Bideshi, D. K., Tan, Y., Spears, T., & Bigot, Y. (2009). Ascoviruses: Superb manipulators of apoptosis for viral replication and transmission. *Current Topics in Microbiology and Immunology*, *328*, 171–196. [https://doi.org/10.1007/978-3-540-68618-7\\_5](https://doi.org/10.1007/978-3-540-68618-7_5)
- Filée, J., & Chandler, M. (2010). Gene exchange and the origin of giant viruses. *Intervirology*, *53*(5), 354–361. <https://doi.org/10.1159/000312920>
- Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Systematic Zoology*, *19*(2), 99–113.
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Galindo, I., & Alonso, C. (2017). African Swine Fever Virus: A Review. *Viruses*, *9*(5), 103. <https://doi.org/10.3390/v9050103>
- Galindo, I., Cuesta-Geijo, M. A., Hlavova, K., Muñoz-Moreno, R., Barrado-Gil, L., Dominguez, J., & Alonso, C. (2015). African swine fever virus infects macrophages, the natural host cells, via clathrin- and cholesterol-dependent endocytosis. *Virus Research*, *200*, 45–55. <https://doi.org/10.1016/j.virusres.2015.01.022>
- Gaudreault, N. N., Madden, D. W., Wilson, W. C., Trujillo, J. D., & Richt, J. A. (2020). African Swine Fever Virus: An Emerging DNA Arbovirus. *Frontiers in Veterinary Science*, *7*, 215. <https://doi.org/10.3389/fvets.2020.00215>

- Geballa-Koukoulas, K., Abdi, S., La Scola, B., Blanc, G., & Andreani, J. (2021). Pacmanvirus S19, the Second Pacmanvirus Isolated from Sewage Waters in Oran, Algeria. *Microbiology Resource Announcements*, *10*(42), e0069321. <https://doi.org/10.1128/MRA.00693-21>
- Geballa-Koukoulas, K., Andreani, J., La Scola, B., & Blanc, G. (2021). The Kaumoebavirus LCC10 Genome Reveals a Unique Gene Strand Bias among “Extended Asfarviridae”. *Viruses*, *13*(2), 148. <https://doi.org/10.3390/v13020148>
- Gelderblom, H. R. (1996). Structure and Classification of Viruses. Em S. Baron (Org.), *Medical Microbiology* (4th ed). University of Texas Medical Branch at Galveston. <http://www.ncbi.nlm.nih.gov/books/NBK8174/>
- Ghigo, E., Kartenbeck, J., Lien, P., Pelkmans, L., Capo, C., Mege, J.-L., & Raoult, D. (2008). Ameobal pathogen mimivirus infects macrophages through phagocytosis. *PLoS Pathogens*, *4*(6), e1000087. <https://doi.org/10.1371/journal.ppat.1000087>
- González-Tortuero, E., Krishnamurthi, R., Allison, H. E., Goodhead, I. B., & James, C. E. (2021). *Comparative analysis of gene prediction tools for viral genome annotation* (p. 2021.12.11.472104). bioRxiv. <https://doi.org/10.1101/2021.12.11.472104>
- Gorbalenya, A. E., & Lauber, C. (2017). Phylogeny of Viruses. *Reference Module in Biomedical Sciences*, B978-0-12-801238-3.95723-4. <https://doi.org/10.1016/B978-0-12-801238-3.95723-4>
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, *57*(Pt 1), 81–91. <https://doi.org/10.1099/ijss.0.64483-0>

- HAJIALILO, E., NIYYATI, M., SOLAYMANI, M., & REZAEIAN, M. (2015). Pathogenic Free-Living Amoebae Isolated From Contact Lenses of Keratitis Patients. *Iranian Journal of Parasitology*, *10*(4), 541–546.
- Hannat, S., La Scola, B., Andreani, J., & Aherfi, S. (2023). Asfarviruses and Closely Related Giant Viruses. *Viruses*, *15*(4), 1015. <https://doi.org/10.3390/v15041015>
- Hernaez, B., Escribano, J. M., & Alonso, C. (2006). Visualization of the African swine fever virus infection in living cells by incorporation into the virus particle of green fluorescent protein-p54 membrane protein chimera. *Virology*, *350*(1), 1–14. <https://doi.org/10.1016/j.virol.2006.01.021>
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. <https://doi.org/10.1093/nar/gky1085>
- Introduction to Virus Taxonomy | ICTV*. ([s.d.]). Recuperado 21 de fevereiro de 2024, de [https://ictv.global/report\\_9th/Introduction/Introduction-to-Virus-Taxonomy](https://ictv.global/report_9th/Introduction/Introduction-to-Virus-Taxonomy)
- Iyer, L. M., Aravind, L., & Koonin, E. V. (2001). Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. *Journal of Virology*, *75*(23), 11720–11734. <https://doi.org/10.1128/JVI.75.23.11720-11734.2001>
- Iyer, L. M., Balaji, S., Koonin, E. V., & Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Research*, *117*(1), 156–184. <https://doi.org/10.1016/j.virusres.2006.01.009>
- Javanmard, E., Niyyati, M., Lorenzo-Morales, J., Lasjerdi, Z., Behniafar, H., & Mirjalali, H. (2017). Molecular identification of waterborne free living amoebae (*Acanthamoeba*, *Naegleria* and *Vermamoeba*) isolated from municipal drinking water and

- environmental sources, Semnan province, north half of Iran. *Experimental Parasitology*, 183, 240–244. <https://doi.org/10.1016/j.exppara.2017.09.016>
- Jouvenet, N., Monaghan, P., Way, M., & Wileman, T. (2004). Transport of African swine fever virus from assembly sites to the plasma membrane is dependent on microtubules and conventional kinesin. *Journal of Virology*, 78(15), 7990–8001. <https://doi.org/10.1128/JVI.78.15.7990-8001.2004>
- Kabra, A., Kabra, R., kukkar, rajiv, Siddiqui, M., & Uddin, K. (2020). *African Swine Fever: An Emerging Viral Disease in India – A Review*. <https://doi.org/10.33980/jbcc.2019.v06i01.003>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), Artigo 6. <https://doi.org/10.1038/nmeth.4285>
- Karki, S., Moniruzzaman, M., & Aylward, F. O. (2021). Comparative Genomics and Environmental Distribution of Large dsDNA Viruses in the Family Asfarviridae. *Frontiers in Microbiology*, 12, 657471. <https://doi.org/10.3389/fmicb.2021.657471>
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kennedy, S., Devine, P., Hurley, C., Ooi, Y.-S., & Collum, L. M. T. (1995). Corneal infection associated with *Hartmannella vermiformis* in contact-lens wearer. *The Lancet*, 346(8975), 637–638. <https://doi.org/10.5555/uri:pii:S0140673695914684>
- Khan, I., Rehman, A., Zia, K., Naveed, U., Bibi, S., Sherazi, R., Hussain, I., Rehman, M. U., & Massa, S. (2020). Microbes and Environment: Global Warming Reverting the Frozen Zombies. Em S. Fahad, M. Hasanuzzaman, M. Alam, H. Ullah, M. Saeed, I. Ali Khan, & M. Adnan (Orgs.), *Environment, Climate, Plant and Vegetation Growth*

- (p. 607–633). Springer International Publishing. [https://doi.org/10.1007/978-3-030-49732-3\\_24](https://doi.org/10.1007/978-3-030-49732-3_24)
- Klemm, P., Stadler, P. F., & Lechner, M. (2023). Proteinortho6: Pseudo-reciprocal best alignment heuristic for graph-based detection of (co-)orthologs. *Frontiers in Bioinformatics*, 3. <https://doi.org/10.3389/fbinf.2023.1322477>
- Klose, T., Reteno, D. G., Benamar, S., Hollerbach, A., Colson, P., La Scola, B., & Rossmann, M. G. (2016). Structure of faustovirus, a large dsDNA virus. *Proceedings of the National Academy of Sciences of the United States of America*, 113(22), 6206–6211. <https://doi.org/10.1073/pnas.1523999113>
- Kolde, R. (2019). *pheatmap: Pretty Heatmaps* (1.0.12) [Software]. <https://cran.r-project.org/web/packages/pheatmap/index.html>
- Konstantinidis, K. T., Rosselló-Móra, R., & Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *The ISME Journal*, 11(11), 2399–2406. <https://doi.org/10.1038/ismej.2017.113>
- Koonin, E. V. (2005). Orthologs, Paralogs, and Evolutionary Genomics. *Annual Review of Genetics*, 39(1), 309–338. <https://doi.org/10.1146/annurev.genet.39.073003.114725>
- Koonin, E. V., Dolja, V. V., Krupovic, M., & Kuhn, J. H. (2021). Viruses Defined by the Position of the Virosphere within the Replicator Space. *Microbiology and Molecular Biology Reviews*, 85(4), e00193-20. <https://doi.org/10.1128/MMBR.00193-20>
- Koonin, E. V., & Yutin, N. (2019). Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. *Advances in Virus Research*, 103, 167–202. <https://doi.org/10.1016/bs.aivir.2018.09.002>
- Krupovic, M., & Koonin, E. V. (2015). Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. *Nature Reviews. Microbiology*, 13(2), 105–115. <https://doi.org/10.1038/nrmicro3389>



- Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278.  
<https://doi.org/10.1093/bioinformatics/btu531>
- Lebreton, F., Manson, A. L., Saavedra, J. T., Straub, T. J., Earl, A. M., & Gilmore, M. S. (2017). Tracing the Enterococci from Paleozoic Origins to the Hospital. *Cell*, *169*(5), 849–861.e13. <https://doi.org/10.1016/j.cell.2017.04.027>
- Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics*, *12*(1), 124. <https://doi.org/10.1186/1471-2105-12-124>
- Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., & Smith, D. B. (2018). Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, *46*(D1), D708–D717.  
<https://doi.org/10.1093/nar/gkx932>
- Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., Lescot, M., Poirot, O., Bertaux, L., Bruley, C., Couté, Y., Rivkina, E., Abergel, C., & Claverie, J.-M. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences*, *111*(11), 4274–4279. <https://doi.org/10.1073/pnas.1320670111>
- Li, G., Fu, D., Zhang, G., Zhao, D., Li, M., Geng, X., Sun, D., Wang, Y., Chen, C., Jiao, P., Cao, L., Guo, Y., & Rao, Z. (2020). Crystal structure of the African swine fever virus structural protein p35 reveals its role for core shell assembly. *Protein & Cell*, *11*(8), 600–605. <https://doi.org/10.1007/s13238-020-00730-w>
- Li, W., Fu, L., Niu, B., Wu, S., & Wooley, J. (2012). Ultrafast clustering algorithms for metagenomic sequence analysis. *Briefings in Bioinformatics*, *13*(6), 656–668.  
<https://doi.org/10.1093/bib/bbs035>

- Liu, S., Luo, Y., Wang, Y., Li, S., Zhao, Z., Bi, Y., Sun, J., Peng, R., Song, H., Zhu, D., Sun, Y., Li, S., Zhang, L., Wang, W., Sun, Y., Qi, J., Yan, J., Shi, Y., Zhang, X., ... Gao, G. F. (2019). Cryo-EM Structure of the African Swine Fever Virus. *Cell Host & Microbe*, 26(6), 836-843.e3. <https://doi.org/10.1016/j.chom.2019.11.004>
- Lopes, I., Altab, G., Raina, P., & de Magalhães, J. P. (2021). Gene Size Matters: An Analysis of Gene Length in the Human Genome. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.559998>
- Lwoff, A. (1957). The concept of virus. *Journal of General Microbiology*, 17(2), 239–253. <https://doi.org/10.1099/00221287-17-2-239>
- Lwoff, A., & Tournier, P. (1966). The classification of viruses. *Annual Review of Microbiology*, 20, 45–74. <https://doi.org/10.1146/annurev.mi.20.100166.000401>
- Matsuyama, T., Kiryu, I., Inada, M., Takano, T., Matsuura, Y., & Kamaishi, T. (2021). Susceptibility of Four Abalone Species, *Haliotis gigantea*, *Haliotis discus discus*, *Haliotis discus hannai* and *Haliotis diversicolor*, to Abalone asfa-like Virus. *Viruses*, 13(11), 2315. <https://doi.org/10.3390/v13112315>
- Matsuyama, T., Kiryu, I., Mekata, T., Takano, T., Umeda, K., & Matsuura, Y. (2023). Pathogenicity, genomic analysis and structure of abalone asfa-like virus: Evidence for classification in the family Asfarviridae. *Journal of General Virology*, 104(8), 001875. <https://doi.org/10.1099/jgv.0.001875>
- Matsuyama, T., Takano, T., Nishiki, I., Fujiwara, A., Kiryu, I., Inada, M., Sakai, T., Terashima, S., Matsuura, Y., Isowa, K., & Nakayasu, C. (2020). A novel Asfarvirus-like virus identified as a potential cause of mass mortality of abalone. *Scientific Reports*, 10(1), 4620. <https://doi.org/10.1038/s41598-020-61492-3>

- Medini, D., Donati, C., Tettelin, H., Massignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, *15*(6), 589–594.  
<https://doi.org/10.1016/j.gde.2005.09.006>
- Meier-Kolthoff, J. P., Klenk, H.-P., & Göker, M. (2014). Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *International Journal of Systematic and Evolutionary Microbiology*, *64*(Pt\_2), 352–356.  
<https://doi.org/10.1099/ij.s.0.056994-0>
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, *19*(6), Artigo 6.  
<https://doi.org/10.1038/s41592-022-01488-1>
- Modrow, S., Falke, D., Truyen, U., & Schätzl, H. (2013). Viruses: Definition, Structure, Classification. *Molecular Virology*, 17–30. [https://doi.org/10.1007/978-3-642-20718-1\\_2](https://doi.org/10.1007/978-3-642-20718-1_2)
- Mönttinen, H. A. M., Bicep, C., Williams, T. A., & Hirt, R. P. (2021). The genomes of nucleocytoplasmic large DNA viruses: Viral evolution writ large. *Microbial Genomics*, *7*(9), 000649. <https://doi.org/10.1099/mgen.0.000649>
- Müller, D. G., Bräutigam, M., & Knippers, R. (1996). Virus infection and persistence of foreign DNA in the marine brown alga *Feldmannia simplex* (Ectocarpales, Phaeophyceae). *Phycologia*, *35*(1), 61–63. <https://doi.org/10.2216/i0031-8884-35-1-61.1>
- Mushegian, A. R. (2020). Are There 1031 Virus Particles on Earth, or More, or Fewer? *Journal of Bacteriology*, *202*(9), e00052-20. <https://doi.org/10.1128/JB.00052-20>
- NAKATSUGAWA, T., HATAI, K., & KUBOTA, S. S. (1988). Histopathological findings on cultured juvenile abalone, *Nordotis discus*, with muscular atrophy. *Fish Pathology*, *23*(3), 203–204. <https://doi.org/10.3147/jsfp.23.203>

- Oliveira, G. P., de Aquino, I. L. M., Luiz, A. P. M. F., & Abrahão, J. S. (2018). Putative Promoter Motif Analyses Reinforce the Evolutionary Relationships Among Faustoviruses, Kaumoebavirus, and Asfarvirus. *Frontiers in Microbiology*, *9*.  
<https://www.frontiersin.org/articles/10.3389/fmicb.2018.01041>
- Padró, J., De Panis, D. N., Luisi, P., Dopazo, H., Szajnman, S., Hasson, E., & Soto, I. M. (2022). Ortholog genes from cactophilic *Drosophila* provide insight into human adaptation to hallucinogenic cacti. *Scientific Reports*, *12*(1), 13180.  
<https://doi.org/10.1038/s41598-022-17118-x>
- Pagnier, I., Valles, C., Raoult, D., & La Scola, B. (2015). Isolation of *Vermamoeba vermiformis* and associated bacteria in hospital water. *Microbial Pathogenesis*, *80*, 14–20. <https://doi.org/10.1016/j.micpath.2015.02.006>
- Penrith, M.-L., Bastos, A. D., Etter, E. M. C., & Beltrán-Alcrudo, D. (2019). Epidemiology of African swine fever in Africa today: Sylvatic cycle versus socio-economic imperatives. *Transboundary and Emerging Diseases*, *66*(2), 672–686.  
<https://doi.org/10.1111/tbed.13117>
- Penrith, M.-L., & Vosloo, W. (2009). Review of African swine fever: Transmission, spread and control. *Journal of the South African Veterinary Association*, *80*(2), 58–62.
- Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J.-M., & Abergel, C. (2013). Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science*, *341*(6143), 281–286.  
<https://doi.org/10.1126/science.1239181>
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, *5*(3), e9490.  
<https://doi.org/10.1371/journal.pone.0009490>

- Queiroz, V. F., Carvalho, J. V. R. P., de Souza, F. G., Lima, M. T., Santos, J. D., Rocha, K. L. S., de Oliveira, D. B., Araújo, J. P., Ullmann, L. S., Rodrigues, R. A. L., & Abrahão, J. S. (2023). Analysis of the Genomic Features and Evolutionary History of Pithovirus-Like Isolates Reveals Two Major Divergent Groups of Viruses. *Journal of Virology*, *97*(7), e00411-23. <https://doi.org/10.1128/jvi.00411-23>
- Quetglas, J. I., Hernáez, B., Galindo, I., Muñoz-Moreno, R., Cuesta-Geijo, M. A., & Alonso, C. (2012). Small rho GTPases and cholesterol biosynthetic pathway intermediates in African swine fever virus infection. *Journal of Virology*, *86*(3), 1758–1767. <https://doi.org/10.1128/JVI.05666-11>
- Rajput, A., Chauhan, S. M., Mohite, O. S., Hyun, J. C., Ardalani, O., Jahn, L. J., Sommer, M. O. A., & Palsson, B. O. (2023). Pangenome analysis reveals the genetic basis for taxonomic classification of the Lactobacillaceae family. *Food Microbiology*, *115*, 104334. <https://doi.org/10.1016/j.fm.2023.104334>
- Rangel-Pineros, G., Almeida, A., Beracochea, M., Sakharova, E., Marz, M., Reyes Muñoz, A., Hölzer, M., & Finn, R. D. (2023). VIRify: An integrated detection, annotation and taxonomic classification pipeline using virus-specific protein profile hidden Markov models. *PLOS Computational Biology*, *19*(8), e1011422. <https://doi.org/10.1371/journal.pcbi.1011422>
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., & Claverie, J.-M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science (New York, N.Y.)*, *306*(5700), 1344–1350. <https://doi.org/10.1126/science.1101485>
- Reteno, D. G., Benamar, S., Khalil, J. B., Andreani, J., Armstrong, N., Klose, T., Rossmann, M., Colson, P., Raoult, D., & La Scola, B. (2015). Faustovirus, an Asfarvirus-Related

- New Lineage of Giant Viruses Infecting Amoebae. *Journal of Virology*, 89(13), 6585–6594. <https://doi.org/10.1128/JVI.00115-15>
- Rigou, S., Santini, S., Abergel, C., Claverie, J.-M., & Legendre, M. (2022). Past and present giant viruses diversity explored through permafrost metagenomics. *Nature Communications*, 13, 5853. <https://doi.org/10.1038/s41467-022-33633-x>
- Rodrigues, R. A. L., Louazani, A. C., Picorelli, A., Oliveira, G. P., Lobo, F. P., Colson, P., La Scola, B., & Abrahão, J. S. (2020). Analysis of a Marseillevirus Transcriptome Reveals Temporal Gene Expression Profile and Host Transcriptional Shift. *Frontiers in Microbiology*, 11. <https://www.frontiersin.org/articles/10.3389/fmicb.2020.00651>
- Rodriguez-R, L. M., & Konstantinidis, K. T. (2016). *The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes* (e1900v1). PeerJ Inc. <https://doi.org/10.7287/peerj.preprints.1900v1>
- Rolland, C., Andreani, J., Cherif Louazani, A., Aherfi, S., Francis, R., Rodrigues, R., Silva, L. S., Sahmi, D., Mougari, S., Chelkha, N., Bekliz, M., Silva, L., Assis, F., Dornas, F., Bou Khalil, J. Y., Pagnier, I., Desnues, C., Levasseur, A., Colson, P., ... La Scola, B. (2019). Discovery and Further Studies on Giant Viruses at the IHU Mediterranean Infection That Modified the Perception of the Virosphere. *Viruses*, 11(4), 312. <https://doi.org/10.3390/v11040312>
- Romiguier, J., Ranwez, V., Douzery, E. J. P., & Galtier, N. (2010). Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. *Genome Research*, 20(8), 1001–1009. <https://doi.org/10.1101/gr.104372.109>
- Rost, B. (1997). Protein structures sustain evolutionary drift. *Folding and Design*, 2, S19–S24. [https://doi.org/10.1016/S1359-0278\(97\)00059-X](https://doi.org/10.1016/S1359-0278(97)00059-X)

- Ruiz Martínez, E., Mckeown, D. A., Schroeder, D. C., Thuestad, G., Sjøtun, K., Sandaa, R.-A., Larsen, A., & Hoell, I. A. (2023). Phaeoviruses Present in Cultured and Natural Kelp Species, *Saccharina latissima* and *Laminaria hyperborea* (Phaeophyceae, Laminariales), in Norway. *Viruses*, *15*(12), Artículo 12.  
<https://doi.org/10.3390/v15122331>
- Sakai, R., & Biederstedt, E. (2021). *dendsort: Modular Leaf Ordering Methods for Dendrogram Nodes* (0.3.4) [Software].  
<https://cran.rstudio.com/web/packages/dendsort/index.html>
- Salas, M. L., & Andrés, G. (2013). African swine fever virus morphogenesis. *Virus Research*, *173*(1), 29–41. <https://doi.org/10.1016/j.virusres.2012.09.016>
- Sánchez, E. G., Quintas, A., Pérez-Núñez, D., Nogal, M., Barroso, S., Carrascosa, Á. L., & Revilla, Y. (2012). African swine fever virus uses macropinocytosis to enter host cells. *PLoS Pathogens*, *8*(6), e1002754. <https://doi.org/10.1371/journal.ppat.1002754>
- Sandaa, R.-A., Heldal, M., Castberg, T., Thyraug, R., & Bratbak, G. (2001). Isolation and Characterization of Two Viruses with Large Genome Size Infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology*, *290*(2), 272–280. <https://doi.org/10.1006/viro.2001.1161>
- Santos-Medellin, C., Zinke, L. A., ter Horst, A. M., Gelardi, D. L., Parikh, S. J., & Emerson, J. B. (2021). Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *The ISME Journal*, *15*(7), Artículo 7.  
<https://doi.org/10.1038/s41396-021-00897-y>
- Schulz, F., Abergel, C., & Woyke, T. (2022). Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nature Reviews Microbiology*, *20*(12), Artículo 12.  
<https://doi.org/10.1038/s41579-022-00754-5>

- Scola, B. L., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.-M., & Raoult, D. (2003). A Giant Virus in Amoebae. *Science*, 299(5615), 2033–2033. <https://doi.org/10.1126/science.1081867>
- Sharma, V., Mobeen, F., & Prakash, T. (2016). Comparative Genomics of Herpesviridae Family to Look for Potential Signatures of Human Infecting Strains. *International Journal of Genomics*, 2016, 9543274. <https://doi.org/10.1155/2016/9543274>
- Siew, N., & Fischer, D. (2003). Analysis of singleton ORFans in fully sequenced microbial genomes. *Proteins*, 53(2), 241–251. <https://doi.org/10.1002/prot.10423>
- Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., Hull, R., King, A. M. Q., Koonin, E. V., Krupovic, M., Kuhn, J. H., Lefkowitz, E. J., Nibert, M. L., Orton, R., Roossinck, M. J., ... Zerbini, F. M. (2017). Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, 15(3), Artigo 3. <https://doi.org/10.1038/nrmicro.2016.177>
- Simmonds, P., Adriaenssens, E. M., Zerbini, F. M., Abrescia, N. G. A., Aiewsakun, P., Alfenas-Zerbini, P., Bao, Y., Barylski, J., Drosten, C., Duffy, S., Duprex, W. P., Dutilh, B. E., Elena, S. F., García, M. L., Junglen, S., Katzourakis, A., Koonin, E. V., Krupovic, M., Kuhn, J. H., ... Vasilakis, N. (2023). Four principles to establish a universal virus taxonomy. *PLOS Biology*, 21(2), e3001922. <https://doi.org/10.1371/journal.pbio.3001922>
- Simões, M., Freitas, F. B., Leitão, A., Martins, C., & Ferreira, F. (2019). African swine fever virus replication events and cell nucleus: New insights and perspectives. *Virus Research*, 270, 197667. <https://doi.org/10.1016/j.virusres.2019.197667>
- Snipen, L., & Liland, K. H. (2015). micropan: An R-package for microbial pan-genomics. *BMC Bioinformatics*, 16(1), 79. <https://doi.org/10.1186/s12859-015-0517-0>



- Sobhy, H. (2017). A comparative review of viral entry and attachment during large and giant dsDNA virus infections. *Archives of Virology*, *162*(12), 3567–3585.  
<https://doi.org/10.1007/s00705-017-3497-8>
- Sun, T.-W., Yang, C.-L., Kao, T.-T., Wang, T.-H., Lai, M.-W., & Ku, C. (2020). Host Range and Coding Potential of Eukaryotic Giant Viruses. *Viruses*, *12*(11), 1337.  
<https://doi.org/10.3390/v12111337>
- Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), Artigo 7057.  
<https://doi.org/10.1038/nature04160>
- Temmam, S., Monteil-Bouchard, S., Sambou, M., Aubadie-Ladrix, M., Azza, S., Decloquement, P., Khalil, J. Y. B., Baudoin, J.-P., Jardot, P., Robert, C., La Scola, B., Mediannikov, O. Y., Raoult, D., & Desnues, C. (2015). Faustovirus-Like Asfarvirus in Hematophagous Biting Midges and Their Vertebrate Hosts. *Frontiers in Microbiology*, *6*. <https://www.frontiersin.org/articles/10.3389/fmicb.2015.01406>
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., DeBoy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., ... Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences*, *102*(39), 13950–13955.  
<https://doi.org/10.1073/pnas.0506758102>
- Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: The bacterial pan-genome. *Current Opinion in Microbiology*, *11*(5), 472–477.  
<https://doi.org/10.1016/j.mib.2008.09.006>

- Thomas, V., Herrera-Rimann, K., Blanc, D. S., & Greub, G. (2006). Biodiversity of amoebae and amoeba-resisting bacteria in a hospital water network. *Applied and Environmental Microbiology*, 72(4), 2428–2438. <https://doi.org/10.1128/AEM.72.4.2428-2438.2006>
- Toll-Riera, M., & Albà, M. M. (2013). Emergence of novel domains in proteins. *BMC Evolutionary Biology*, 13, 47. <https://doi.org/10.1186/1471-2148-13-47>
- van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J., & Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, 42(2), 243–246. <https://doi.org/10.1038/s41587-023-01773-0>
- Venter, S. N., Palmer, M., & Steenkamp, E. T. (2022). Relevance of prokaryotic subspecies in the age of genomics. *New Microbes and New Infections*, 48, 101024. <https://doi.org/10.1016/j.nmni.2022.101024>
- Vernikos, G. S. (2020). A Review of Pangenome Tools and Recent Studies. Em H. Tettelin & D. Medini (Orgs.), *The Pangenome: Diversity, Dynamics and Evolution of Genomes*. Springer. <http://www.ncbi.nlm.nih.gov/books/NBK558826/>
- Vogel Ely, C., Bordignon, S. A. de L., Trevisan, R., & Boldrini, I. I. (2017). Implications of poor taxonomy in conservation. *Journal for Nature Conservation*, 36, 10–13. <https://doi.org/10.1016/j.jnc.2017.01.003>
- Wang, L., Luo, Y., Zhao, Y., Gao, G. F., Bi, Y., & Qiu, H. (2020). Comparative genomic analysis reveals an ‘open’ pan-genome of African swine fever virus. *Transboundary and Emerging Diseases*, 67(4), 1553–1562. <https://doi.org/10.1111/tbed.13489>
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40(7), e49. <https://doi.org/10.1093/nar/gkr1293>

- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Brand, T. van den, Posit, & PBC. (2024). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (3.5.1) [Software]. <https://cran.r-project.org/web/packages/ggplot2/index.html>
- Willger, S. D., Cornish, E. J., Chung, D., Fleming, B. A., Lehmann, M. M., Puttikamonkul, S., & Cramer, R. A. (2012). Dsc orthologs are required for hypoxia adaptation, triazole drug responses, and fungal virulence in *Aspergillus fumigatus*. *Eukaryotic Cell*, *11*(12), 1557–1567. <https://doi.org/10.1128/EC.00252-12>
- Wilson, W. H., Van Etten, J. L., & Allen, M. J. (2009). The Phycodnaviridae: The Story of How Tiny Giants Rule the World. *Current topics in microbiology and immunology*, *328*, 1–42.
- Witt, A. S. A., Carvalho, J. V. R. P., Serafim, M. S. M., Arias, N. E. C., Rodrigues, R. A. L., & Abrahão, J. S. (2024). *The GC% landscape of the Nucleocyotoviricota*. <https://doi.org/10.21203/rs.3.rs-4093445/v1>
- Yu, Z., Zhang, W., Fu, H., Zou, X., Zhao, M., Liang, S., Gu, C., Yang, Q., He, M., Xiao, Q., Xiao, W., He, L., & Lü, M. (2021). Genomic analysis of Poxviridae and exploring qualified gene sequences for phylogenetics. *Computational and Structural Biotechnology Journal*, *19*, 5479–5486. <https://doi.org/10.1016/j.csbj.2021.09.031>
- Yuan, W.-G., Liu, G.-F., Shi, Y.-H., Xie, K.-M., Jiang, J.-Z., & Yuan, L.-H. (2022). A discussion of RNA virus taxonomy based on the 2020 International Committee on Taxonomy of Viruses report. *Frontiers in Microbiology*, *13*, 960465. <https://doi.org/10.3389/fmicb.2022.960465>
- Yutin, N., Wolf, Y. I., & Koonin, E. V. (2014). Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology*, *466–467*, 38–52. <https://doi.org/10.1016/j.virol.2014.06.032>

Yutin, N., Wolf, Y. I., Raoult, D., & Koonin, E. V. (2009). Eukaryotic large nucleocytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution. *Virology Journal*, 6, 223. <https://doi.org/10.1186/1743-422X-6-223>

## ANEXO A - PRODUÇÃO ACADÊMICA

Nesta seção são apresentados a produção acadêmica durante o período do doutorado do autor (2020-2024).

Artigos publicados como primeiro autor:

**SANTOS, Thiago M.** et al. Human variation in the protein receptor ACE2 affects its binding affinity to SARS-CoV-2 in a variant-dependent manner. *Journal of Biomolecular Structure and Dynamics*, v. 41, n. 7, p. 2947-2955, 2023.

SOUZA, J. S. M; A B P LISBOA; **SANTOS, Thiago M.**, et al. The evolution of ADAM gene family in eukaryotes. *Genomics*, v. 112, n. 5, p. 3108-3116, 2020.

**SANTOS, Thiago M.** et al. Characterization and Phylogenetic Insights into the First Viral Major Vault Proteins (MVPs), Identified in Bacteriophages from the Human Gut. *bioRxiv*, p. 2023.11. 10.566534, 2023.

Demais artigos:

MACHADO, Talita B.; Picorelli ACR; de Azevedo BL; de Aquino ILM; Queiroz VF; Rodrigues RAL; Araújo JP Jr; Ullmann LS; **Dos Santos TM**; Marques RE; Guimarães SL; Andrade ACSP; Gualarte JS; Demoliner M; Filippi M; Pereira VMAG; Spilki FR; Krupovic M; Aylward FO; Del-Bem L-E; Abrahão JS. Gene duplication as a major force driving the genome expansion in some giant viruses. *Journal of Virology*, v. 97, n. 12, p. e01309-23, 2023.

Livros/capítulos de livros:

**SANTOS, Thiago M.**; Menezes, A.P.A.; Rodrigues, W.F.C.; Del Bem, L.E.. Vigilância Genômica em Saúde Pública. *Vigilância em Saúde: Interfaces entre a Saúde Pública e a Pesquisa Científica*. 1ed.: Editora COLAB, 2021, v., p. 32-50