

FUNDAÇÃO OSWALDO CRUZ
INSTITUTO AGGEU MAGALHÃES
Mestrado Acadêmico em Biociências e Biotecnologia em Saúde

FILIFE ZIMMER DEZORDI

**CARACTERIZAÇÃO DE VÍRUS INSERIDOS NOS GENOMAS DE
CULICÍDEOS VETORES DE PATÓGENOS**

RECIFE

2020

FILIPPE ZIMMER DEZORDI

**CARACTERIZAÇÃO DE VÍRUS INSERIDOS NOS GENOMAS DE
CULICÍDEOS VETORES DE PATÓGENOS**

Dissertação apresentada ao Mestrado em
Biociências e Biotecnologia em Saúde do
Instituto Aggeu Magalhães, Fundação Oswaldo
Cruz, como requisito para a obtenção do título
de Mestre em Ciências.

Orientador: Dr. Gabriel da Luz Wallau

RECIFE

2020

Catálogo na fonte: Biblioteca do Instituto Aggeu Magalhães

- D532c Dezordi, Filipe Zimmer.
Caracterização de vírus inseridos nos genomas de culicídeos vetores de patógenos / Filipe Zimmer Dezordi. Recife: [s.n.], 2020.
94 p. : il., graf., tab. ; 30 cm.
- Dissertação (Mestrado Acadêmico em Biociências e Biotecnologia em Saúde) - Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2020.
Orientador: Gabriel da Luz Wallau.
1. Culicidae - genética. 2. Culicidae - fisiologia. 3. Genoma de insetos. 4. Vírus - genética. 5. Insetos vetores. 6 Genes Virais. 7. Filogenia. 8. Variação Genética. 9. Evolução molecular. 10. Interações Hospedeiro-Patógeno - virologia. I. Wallau, Gabriel da Luz. II. Título.

CDU 595.771

FILIPPE ZIMMER DEZORDI

**CARACTERIZAÇÃO DE VÍRUS INSERIDOS NOS GENOMAS DE
CULICÍDEOS VETORES DE PATÓGENOS**

Dissertação apresentada ao Mestrado em
Biociências e Biotecnologia em Saúde do
Instituto Aggeu Magalhães, Fundação Oswaldo
Cruz, como requisito para a obtenção do título
de Mestre em Ciências.

Aprovado em: 19/02/2020

BANCA EXAMINADORA

Dr. Gabriel da Luz Wallau (Orientador)
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz

Dr. Roberto Dias Lins Neto (Titular interno)
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz

Dr. Luiz Gustavo Rodrigues Souza (Titular externo)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Aos meus pais, Gilmar e Angelita, por darem sentido a tudo isso.

Ao meu irmão Vitchenzo, pelo exemplo de obstinação.

Ao meu orientador Gabriel, pela amizade e pelo exemplo de profissionalismo.

Aos amigos Alexandre e Laís, *Ohana*.

À Luísa, pelo companheirismo e pelo apoio ao final dessa jornada, e durante início da próxima.

Aos amigos Aline, Larissa e Jefferson, que *do nada*, tornaram esse processo muito mais agradável.

Ao meu *homemate* Bruno, pelas conversas, bolos e filosofias sobre a vida.

Aos amigos Dennis, França, Helber, Hudson, Lucas, Luciano, Ludgero, Natália e Thamires, que apesar da distância, não deixaram a amizade enfraquecer ao longo desses dois anos.

A toda equipe do núcleo de bioinformática, em especial à Crhisllane, Elverson e João. Foi durante conversas, *journals* e pedidos de auxílio à vocês que parei de “catar milhos na tela preta” e comecei a desenvolver meus próprios *scripts*.

Por último, mas não menos importante, a todas as pessoas que mantêm o funcionamento da Instituto Aggeu Magalhães e das agências de fomento CAPES e FACEPE, e a todos que cruzaram algum momento dessa jornada comigo, em um período no qual estudantes de pós graduação possuem seis vezes mais chances de desenvolver depressão e ansiedade, posso dizer foi um prazer desenvolver meu projeto nessa instituição, e mesmo estando a mais de 4 mil Km da minha família, pude me sentir em casa, e aparentemente finalizar o mestrado com minha saúde mental “em dia”.

*Nothing in Biology Makes Sense
except in the Light of Evolution*

Theodosius Dobzhansky

DEZORDI, Filipe Zimmer. **Caracterização de vírus inseridos nos genomas de culicídeos vetores de patógenos**. 2020. Dissertação (Mestrado Acadêmico em Biociências e Biotecnologia em Saúde) - Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2020.

RESUMO

Elementos virais endogenizados (EVEs) são sequências de origem viral que se integraram no genoma do seu hospedeiro. Nos últimos anos, uma quantidade crescente de estudos tem demonstrado que EVEs podem adquirir uma variedade de novas funções, modificando ou originando novos genes e redes regulatórias. Mosquitos são artrópodes vetores de vários arbovírus (vírus transmitidos por Artrópodes) responsáveis por causar infecções nas populações humanas: como o vírus da dengue, chikungunya e Zika. Arbovírus possuem uma forte associação com algumas espécies de mosquitos, o que sugere que estes tenham co-evoluído com culicídeos há alguns milhares de anos. A disponibilidade de dados genômicos de culicídeos nos permite compreender as relações evolutivas e o impacto dos EVEs no genoma desses insetos. Dessa maneira, este projeto teve como objetivo caracterizar a diversidade de EVEs presentes nos genomas de 36 espécies de culicídeos, os quais representam, em sua maioria, espécies de importância médica/epidemiológica, bem como identificar possíveis impactos nestes genomas. Foram identificadas 1998 sequências de origem viral no material genético de 36 espécies de culicídeos, representando um total de 27 famílias virais. As famílias *Rhabdoviridae*, *Chuviridae* e *Flaviviridae* apresentam a maior quantidade de elementos endogenizados, com 597, 459 e 228 elementos, respectivamente. Foi possível relacionar dois possíveis eventos evolutivos atrelados a EVEs da família *Chuviridae*, a possível origem de um novo retrovírus, fruto da hibridização de retroelementos Pao com o gene que é traduzido em glicoproteína de Chuvirus, e um possível mecanismo antiviral a partir de glicoproteínas endogenizadas. Estes resultados refletem uma grande diversidade de vírus integrados ao genoma de culicídeos, e levantam questões sobre a importância desses elementos nos genomas estudados.

Palavras chave: Vírus. Mosquito. Evolução. Vetor.

DEZORDI, Filipe Zimmer. **Characterization of viruses inserted in the genomes of culicidae vectors of pathogens**. 2020. Dissertation (Master's Degree in Biosciences and Biotechnology in Health) - Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2020.

ABSTRACT

Endogenous Viral Elements (EVEs) are viral-derived sequences that became integrated into the host genome. In the last few years, several studies have demonstrated that EVEs can give origin to new functions at the molecular level, modifying or creating new genes and regulatory networks. Mosquitoes are arthropods vectors of several Arboviruses (*Arthropod-borne viruses*), the causative agents of human diseases like: Dengue, Chikungunya and Zika fever. Arboviruses have a strong association with mosquito species suggesting that the co-evolutionary process between mosquitoes and arboviruses is occurring since thousand years ago. The genomic data from culicids currently available allow us to investigate the evolutionary impact of EVEs using a comparative genomics approach. In this way, the main goal of this project is characterize the endogenous viral elements diversity on genomes from 36 species of mosquitoes, which represents, in most cases, species with medical and epidemiological importance, as well to identify the potential impacts on the host genome, by using public data and data generated in collaborative projects of the Instituto Aggeu Magalhães. Our results revealed 1998 EVEs sequences found on the genetic material of 36 mosquito species, representing a total of 27 viral families. The major part of endogenous elements belongs to the families *Rhabdoviridae*, *Chuviridae* and *Flaviviridae* with 597, 459, 228 EVEs respectively. Moreover, we also identified two evolutionary events related to EVEs from *Chuviridae* family: the putative origin of a new retrovirus due the hybridization of Pao retroelements and the Env gene of Chuviruses, and a putative antiviral mechanism from endogenized glycoproteins. Those results bring to light a wide diversity of viruses integrated into the mosquito genomes that infected or currently infect culicids and put on perspective which is the importance of these elements in the genomes studied.

Keywords: Viruses. Mosquitoes. Evolution. Vector.

LISTA DE ILUSTRAÇÕES

Figura 1-	Campos relacionados ao surgimento de epidemias de arboviroses transmitidas por culicídeos	18
Figura 2-	Sistema de classificação viral de Baltimore e sistema taxonômico proposto pelo ICTV	22
Figura 3-	Representação esquemática da estrutura genômica de representantes virais já relatados infectando mosquitos	25
Figura 4-	Representação do panorama evolutivo entre os organismos	32
Figura 5-	Representação amostral de gerações filogenéticas geradas em inferência bayesiana	40
Figura 6-	Desenho experimental	45
Figura 7-	Estratégia de desenho de primers para os vírus endógenos	53
Figura 8-	Diversidade de famílias virais, estrutura genômica e hospedeiros de origem de EVEs identificados em genomas de culicídeos	55
Figura 9-	Diversidade de elementos por espécies de culicídeo e comparação com dados da literatura	57
Figura 10-	Distribuição de EVEs por espécie/linhagem de culicídeo	58
Figura 11-	Distribuição de EVEs por grupo de proteínas endogenizadas	60
Figura 12-	Exemplos de glicoproteínas de Chuvirus endogenizados em diferentes contextos genômicos	63
Figura 13-	Modelagem das estruturas tridimensionais das glicoproteínas	65
Figura 14-	Representação filogenética de glicoproteínas de chuvirus endogenizados	67

Figura 15-	Representação filogenética das regiões de Transcriptase Reversa e RNase H de poliproteínas de BEL-Pao	68
Figura 16-	Panorama dos EVEs derivados de Chuvirus	69
Figura 17-	Panorama evolutivo atual de elementos da superfamília BEL/Pao	74

LISTA DE TABELAS

Tabela 1-	Genomas de Culicídeos disponíveis nos bancos de dados públicos: NCBI e VectorBase	16
Tabela 2-	Genomas de Culicídeos gerados em projetos colaborativos no Instituto Aggeu Magalhães	47
Tabela 3-	Glicoproteínas submetidas à processos de modelagem molecular	64
Tabela 4-	Elementos validados por PCR	70

LISTA DE ABREVIATURAS E SIGLAS

CHIKV - *Chikungunya virus*

DENV - *Dengue virus*

DNA - *Deoxyribonucleic Acid*

dsDNA - *Double strand Deoxyribonucleic Acid*

dsRNA - *Double strand Ribonucleic Acid*

enJSRV - *Endogenous Jaagsiekte sheep retrovirus*

ERV - *Endogenous Retrovirus*

EVES - *Elementos Virais Endogenizados*

ICTV - *International Committee on Taxonomy of Viruses*

iRNA - *Ribonucleic Acid Interference*

ISV - *Insect Specific Virus*

JSRV - *Jaagsiekte sheep retrovirus*

kb - *quilo base*

kbp - *quilo base pair*

NCBI - *National Center for Biotechnology Information*

PCR - *Polymerase Chain Reaction*

piRNA - *PIWI-interacting Ribonucleic Acid*

RNA - *Ribonucleic Acid*

ssDNA - *Single strand Deoxyribonucleic Acid*

ssDNA-RT - *Single strand Deoxyribonucleic Acid with Reverse Transcription*

sRNA - *Small Ribonucleic Acid*

ssRNA-RT - *Single strand Ribonucleic Acid with Reverse Transcription*

YFV - *Yellow fever virus*

ZIKV - *Zika virus*

(+)ssRNA - *Single strand Ribonucleic Acid positive sense*

(-)ssRNA - *Single strand Ribonucleic Acid negative sense*

SUMÁRIO

1 INTRODUÇÃO	13
2 REFERENCIAL TEÓRICO	15
2.1 Culicídeos: Biologia Geral	15
2.2 Vírus e Arbovírus	19
<u>2.2.1 Taxonomia</u>	20
<u>2.2.2 Estrutura Genômica</u>	23
2.2.2.1 Ordem Muvirales	24
2.2.2.2 Ordem Jingchuvirales	24
2.2.2.3 Ordem Mononegavirales	24
2.2.2.4 Ordem Bunyavirales	26
2.2.2.5 Ordem Articulavirales	26
2.2.2.6 Ordem Herpesvirales	27
2.2.2.7 Ordem Nidovirales	30
2.2.2.8 Ordem Ortervirales	30
2.2.2.9 Ordem Picornavirales	28
2.2.2.10 Família Flaviviridae	30
2.2.2.11 Família Reoviridae	30
2.2.2.12 Família Togaviridae	30
2.2.2.13 Família Parvoviridae	30
<u>2.2.3 Relações Evolutivas</u>	31
2.3 Elementos virais endogenizados	32
2.4 Paleovirologia	35
2.5 Importância de EVEs em análises de Metavirômica	37
2.6 Bioinformática como ferramenta de análise de vírus endogenizados	37
<u>2.6.1 Análises filogenéticas: reconstruindo a história evolutiva de sequências genômicas</u>	38
<u>2.6.2 Caracterização estrutural</u>	39
3 JUSTIFICATIVA	41
4 PERGUNTA CONDUTORA	42
5 HIPÓTESE	43
6 OBJETIVOS	44
6.1 Objetivo Geral	44
6.2 Objetivos Específicos	44

7 METODOLOGIA	45
7.1 Obtenção de dados	46
<u>7.1.1 Datasets privados</u>	47
<u>7.1.2 Genomas disponíveis em bancos públicos</u>	47
7.2 Identificação de potenciais EVEs	48
<u>7.2.1 Removendo falsos positivos</u>	49
<u>7.2.2 Classificação Viral</u>	49
<u>7.2.3 Obtenção de sequências proteicas preditas</u>	50
7.3 Predição e função de contexto genômico	50
<u>7.3.1 Estrutura das regiões flanqueadoras</u>	50
7.3.1.1 Transposons	50
7.3.1.2 Genes	51
<u>7.3.2 Alinhamento e obtenção de blocos conservados</u>	51
<u>7.3.3 Reconstruções evolutivas</u>	51
<u>7.3.4 Modelagem molecular</u>	52
7.4 Análises in vitro	52
7.5 Visualização de dados	53
8 RESULTADOS	54
8.1 Diversidade de EVEs em culicídeos: Origem de hospedeiros, distribuição em culicídeos e regiões endogenizadas	54
8.2 Família Chuviridae	59
<u>8.2.1 Análises de regiões flanqueadoras:</u>	61
<u>8.2.2 Modelagem molecular</u>	62
<u>8.2.3 Análises Evolutivas</u>	66
<u>8.2.4 Validações das solo-glicoproteínas em bancada:</u>	69
9 DISCUSSÃO	71
10 CONCLUSÃO	75
11 CRONOGRAMA DE EXECUÇÃO	76
REFERÊNCIAS	77
APÊNDICE - INFORMAÇÕES ADICIONAIS	90

1 INTRODUÇÃO

Desde a revolução genômica, que teve seu início nas décadas de 80 e 90, muito vem sendo feito para caracterizar a composição codificante e não codificante dos genomas e entender os processos evolutivos que as modificam. A caracterização de genomas revelou padrões inesperados, como a presença massiva de parasitas genômicos, conhecidos como Elementos de Transposição, ou transposons, e Elementos Virais Endogenizados (EVEs), na grande maioria das espécies eucarióticas (KIDWELL; LISCH, 2000; KOONIN; DOLJA; KRUPOVIC, 2015). Esses elementos, na maioria das vezes, não contribuem para o valor adaptativo dos organismos, ao contrário, esses parasitas utilizam a maquinaria molecular do organismo hospedeiro para promover sua própria propagação. Entretanto, alguns trabalhos, já têm mostrado que EVEs podem trazer modificações genéticas úteis para o genoma hospedeiro, como mudanças na regulação gênica e respostas antivirais (ARMEZZANI *et al.*, 2014; FESCHOTTE; GILBERT, 2012).

As propriedades infectivas de vírus permitem que estes sejam transferidos horizontalmente (transferência gênica entre organismos sem vínculo parental) entre indivíduos da mesma espécie ou de espécies diferentes. Entretanto, diversos vírus podem se tornar parte do material genético da sua espécie hospedeira, um fenômeno conhecido como endogenização. Uma vez endogenizados, os vírus são transmitidos para todos os descendentes do organismo parasitado através do processo de transferência vertical (transferência gênica de um organismo para sua prole) (FESCHOTTE; GILBERT, 2012), um processo que ocorre em praticamente todos os domínios da vida, em procariotos, arqueas e eucariotos. Dentre os eucariotos, grande parte do conhecimento relacionado a EVEs se restringe a diferentes espécies de vertebrados, apesar de um número crescente de achados indicar que EVEs também podem ser facilmente encontrados em invertebrados incluindo insetos (TER HORST *et al.*, 2019).

Culicídeos, em sua grande maioria, são insetos vetores para uma gama de vírus patogênicos à humanos como Dengue, Chikungunya, Zika, Febre Amarela entre outros. Alguns trabalhos recentes identificaram genes virais da família *Flaviviridae* (a mesma família a qual pertencem o vírus da Dengue, Zika e Febre Amarela) tanto no genoma de *Aedes aegypti* como *Aedes albopictus* (CHEN, XIAO-GUANG *et al.*, 2015; FORT *et al.*, 2012). Entretanto, esses trabalhos focaram em alguns grupos virais específicos e somente em um ou poucos genomas de culicídeos vetores. Dessa maneira, uma análise detalhada de todos

os grupos virais utilizando uma abordagem de genômica comparativa ainda se faz necessária para compreender completamente os EVEs presentes nos genomas de mosquitos vetores.

Até o momento de início deste trabalho 37 genomas de culicídeos vetores estavam disponíveis em bancos de dados públicos NCBI (do Inglês *National Center for Biotechnology Information*) (NCBI, 2018) e VectorBase (VECTORBASE, 2018). Essa grande quantidade de dados abrangendo genomas de culicídeos que se separaram desde 190 milhões de anos (LEACHE, 2010) até espécies crípticas como *Anopheles gambiae* e *Anopheles arabiensis*, que ainda apresentam sinal de introgressão (BESANSKY *et al.*, 2003), podem fornecer dados valiosos para avaliação de vírus endogenizados e seus padrões evolutivos.

Tomando em conjunto estes dados, o presente estudo teve como objetivo caracterizar a diversidade de EVEs presentes nos genomas de Culicídeos, bem como identificar possíveis impactos nestes genomas, usando dados disponíveis em bancos públicos e dados gerados a partir de projetos em desenvolvimento no Instituto Aggeu Magalhães durante os anos de 2017 e 2018.

2 REFERENCIAL TEÓRICO

Alguns conceitos fundamentais sobre a biologia dos culicídeos e dos vírus devem ser considerados com atenção para o entendimento das relações evolutivas destes organismos.

2.1 Culicídeos: Biologia Geral

Culicídeos são insetos taxonomicamente agrupados na família *Culicidae* (LOEW; MEIGEN, 1817), sendo então subdivididos em 2 subfamílias, *Anophelinae*, a qual engloba todas as espécies do gênero *Anopheles* (GRASSI, 1900) e *Culicinae*, a qual engloba todos os demais gêneros, como por exemplo, *Aedes*, *Culex*, *Psorophora* e *Aedeomyia* (LOEW; MEIGEN, 1817). Das 3.568 espécies descritas até agosto de 2019 (MOSQUITO TAXONOMY INVENTORY, 2019) 27 espécies de três gêneros (*Anopheles*, *Culex* e *Aedes*) possuem seu genoma completo disponível em bancos públicos, totalizando 37 genomas quando consideradas diferentes linhagens da mesma espécie (Tabela 1).

Com exceção do gênero *Toxorhynchites*, todas as espécies da família *Culicidae* possuem hábitos alimentares hematofágicos (CONSOLI; OLIVEIRA, 1994). A alimentação sanguínea conhecida como repasto sanguíneo é um processo necessário para obtenção de fonte energética e fonte de ferro, um elemento essencial para a maturação de ovos (CLEMENTS, 1992). A fonte para o repasto sanguíneo é amplamente variável, podendo ser animais endotérmicos, ou de sangue quente, como mamíferos e aves, quanto animais ectotérmicos, ou de sangue frio, como répteis e anfíbios, não havendo um fator dominante para a escolha da fonte alimentar, e sim uma série de fatores e condições que influenciam a fêmea a procurar determinadas espécies, como por exemplo: oferta alimentar, clima e variação genética do mosquito (TAKKEN; VERHULST, 2013).

Considerando a hematofagia como um processo essencial para o ciclo de vida da maioria das espécies de mosquitos, e que estes se alimentam de uma gama mamíferos incluindo primatas não-humanos e humanos. Os primeiros estudos relatando potenciais espécies antropofílicas datam do século XVII e XVIII, onde populações de mosquitos foram comumente encontradas em ambientes com populações humanas (REDI, 1688; REAMUR, 1740 *apud* FORATTINI, 2002, p. 27).

Nos séculos XIX e XX, com o desenvolvimento de trabalhos na área de entomologia médica, surgiram os primeiros estudos apontando os mosquitos como potenciais agentes transmissores de patógenos, ou seja, organismos que podem atuar como agente de

transferência de patógenos entre hospedeiros, como transmissor da filariose bancroftiana (MANSON, 1879 *apud* FORATTINI, 2002, p. 27) e da Febre Amarela (FINLAY, 1937), por exemplo.

Dentre os diversos tipos de patógenos que podem ser carreados por culicídeos (TOLLE, 2009), vírus tem se sobressaído quando a temática envolve surtos epidêmicos mundiais (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2018), onde, em menos de duas décadas, doenças como Febre Amarela causada por YFV (do Inglês *Yellow Fever Virus*), Dengue causada por DENV (do Inglês *Dengue Virus*), Chikungunya causada por CHIKV (do Inglês *Chikungunya Virus*) e mais recentemente Zika causada por ZIKV (do Inglês *Zika Virus*) emergiram acometendo milhares de seres humanos (BULIVA *et al.*, 2017) e primatas não-humanos, como no caso da Febre Amarela (BARRETT, 2018; MASSAD; MIGUEL; COUTINHO, 2018).

Tabela 1 - Genomas de Culicídeos disponíveis nos bancos de dados públicos: NCBI e VectorBase

(continua)

Espécie e linhagem	Código ¹	Tamanho do Genoma (Mb) ²	Nº. de Scaffolds ³	N50 ⁴
<i>Aedes aegypti</i> AaegL3	Aae_liv	1.311.011.677	4.757	1.547.048
<i>Aedes aegypti</i> Aag2	Aae_Aag2	1.720.000.000	3.752	1.400.000
<i>Aedes aegypti</i> BV	Aae_BV	744.596.036	223.039	4251
<i>Aedes albopictus</i> Rimini	Aal_Rim	1.341.044.778	3.342.920	397
<i>Aedes albopictus</i> Foshan	Aal_Fosh	1.923.476.627	154.782	201.017
<i>Aedes albopictus</i> C636	Aal_C636	2.247.289.552	2.434	3.303.944
<i>Culex quinquefasciatus</i>	CQui_Joh	579.057.705	3.172	486.756
<i>Anopheles albimanus</i>	An_al_ST	173.339.240	236	9.735.467
<i>Anopheles arabiensis</i>	An_ar_Don	246.567.867	1.214	5.604.218
<i>Anopheles atroparvus</i>	An_atr_EB	224.290.125	1.371	9.206.694
<i>Anopheles christyi</i>	An_chr_AC	172.658.580	30.369	9.057
<i>Anopheles coluzzi</i>	An_col_Ma	224.455.335	10.521	10.521
<i>Anopheles culicifacies</i>	An_cul_A	202.998.806	16.162	22.320
<i>Anopheles darlingi</i>	An_dar_Co	136.935.538	2.221	115.072
<i>Anopheles dirus</i>	An_dir_WR	216.307.690	1.266	6.906.475
<i>Anopheles epiroticus</i>	An_epi_E	223.486.714	2.673	366.526
<i>Anopheles farauti</i>	An_far_F	183.103.254	310	12.895.223

Tabela 1 - Genomas de Culicídeos disponíveis nos bancos de dados públicos: NCBI e VectorBase

(conclusão)

Espécie e linhagem	Código¹	Tamanho do Genoma (Mb)²	Nº. de Scaffolds³	N50⁴
<i>Anopheles funestus</i>	An_fun_F	225.223.604	1.392	671.960
<i>Anopheles gambiae PEST</i>	An_gam_PE	273.109.044	8	49 364 325
<i>Anopheles gambiae Pimpirena</i>	An_gam_Pim	293.748.066	8000	101465
<i>Anopheles gambiae S</i>	An_gam_S	236.403.076	13.042	3.800.922
<i>Anopheles gambiae BV</i>	An_gam_BV	220.394.761	54.385	6406
<i>Anopheles koliensis</i>	An_kol_AK	151.110.088	41.925	4.659
<i>Anopheles maculatus m3</i>	An_mac	141.894.015	47.797	3.841
<i>Anopheles maculatus BtQ1</i>	An_mac_BtQ1	462.688.933	50.590	198.528
<i>Anopheles melas</i>	An_mel_CM	224.162.116	20.229	18.103
<i>Anopheles merus</i>	An_mer_MA	288.048.996	2.027	1.489.982
<i>Anopheles minimus</i>	An_min	201.793.324	678	10.313.149
<i>Anopheles nili</i>	An_nil_Din	98.320.046	51.048	2.577
<i>Anopheles punctulatus</i>	An_pun_AP	146.157.495	20.774	10.256
<i>Anopheles quadriannulatus</i>	An_qua	283.828.998	2823	1.641.272
<i>Anopheles sinensis SINENSIS</i>	An_sin_SIN	375.763.635	10.448	579.086
<i>Anopheles sinensis china</i>	An_sin_CH	220.777.669	9.592	814.231
<i>Anopheles stephensi Indian</i>	An_ste_IN	221.324.304	23.371	1.591.355
<i>Anopheles stephensi SDA</i>	An_ste_SD	225.369.006	1.100	837.295
<i>Anopheles cracens</i>	An_cra	326.465.827	34.365	153.896
<i>Anopheles aquasalis</i>	An_aqu	162.944.031	16.504	53.365

Fonte: O autor.

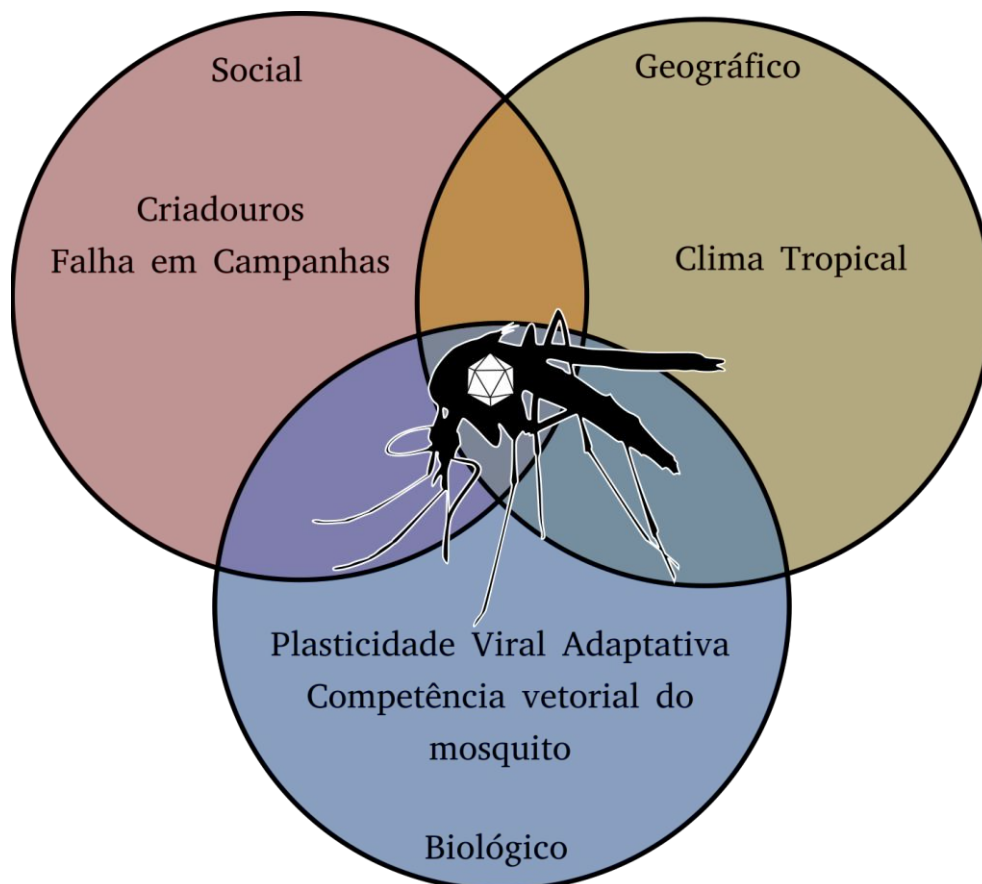
Nota: ¹Código gerado para o presente estudo; ² em Mega bases; ³ sequência formada pela sobreposição de contigs; ⁴ Métrica que representa o tamanho do menor contig em um conjunto de contigs que compõem 50% da montagem.

O aumento de surtos e epidemias envolvendo arbovírus (do Inglês *Arthropod-borne Virus*) pode ser atribuído a fatores de diferentes campos (Figura 1), como por exemplo: No campo social têm-se a criação de habitats urbanos que favorecem a reprodução do mosquito vetor e a falha de campanhas de combate ao vetor em larga escala (GUBLER, 2004). Com relação ao campo geográfico, fatores como o clima impactam diretamente no sucesso de colonização de diversas espécies de culicídeos (ASAD; CARPENTER, 2018; CAMINADE

et al., 2014), sendo então as regiões tropicais as mais afetadas por superpopulações de culicídeos (LAFFERTY; MORDECAI, 2016).

No campo biológico, têm-se fatores como: a plasticidade adaptativa de certos grupos virais - que possuem a capacidade de infectar diversas espécies de mosquitos - bem como espécies de hospedeiros vertebrados, que podem atuar como hospedeiro final ou reservatório viral (WEAVER; BARRETT, 2004; WEBSTER; BORLASE; RUDGE, 2017); e a taxa evolutiva de alguns grupos virais, como vírus de RNA por exemplo, que possuem uma taxa mutacional mais elevada que de outros vírus (SANJUÁN *et al.*, 2010); (PFEIFFER; KIRKEGAARD, 2005; VIGNUZZI *et al.*, 2005), como é o caso dos vírus DENV, YFV, CHIKV e ZIKV envolvidos em epidemias recentes.

Figura 1 - Campos relacionados ao surgimento de epidemias de arboviroses transmitidas por culicídeos



Fonte: O autor, baseado em Gubler (2004); Asad e Carpenter (2018); Caminade *et al.* (2014); Weaver e Barrett (2004); Webster, Borlase e Rudge (2017); Sanjuán *et al.* (2010).

Dois fatores são mensurados para determinar se um organismo até então não considerado vetor, ou considerado um vetor em potencial, é um real vetor biológico: a competência e a capacidade vetorial. A competência vetorial pode ser descrita como a probabilidade de um organismo se infectar com o patógeno e permitir que o mesmo complete seu ciclo de replicação, o que engloba a transmissão para outro hospedeiro (RODHAIN; ROSEN, 1997 *apud* BRAGA, *et al.*, 2004). A capacidade vetorial engloba a competência vetorial juntamente com fatores externos ao organismo potencialmente vetor, como por exemplo: fatores ambientais (clima), populacionais (abundância vetor/hospedeiro) e comportamentais (tropismos) (FORATTINI, 1992).

Além do nível de espécie, a capacidade vetorial pode variar a nível de subpopulações da mesma espécie, onde por exemplo, mosquitos da espécie *Aedes albopictus* mostraram diferenças significativas de competência vetorial quando comparadas populações das Américas do Sul, Central e do Norte (CHOUIN-CARNEIRO *et al.*, 2016). Sabe-se que populações da mesma espécie sofrem diferentes pressões evolutivas exercidas pelas condições bióticas e abióticas (DARWIN, 1859) e que isto reflete em diferenças à nível molecular, estudos têm sugerido que estas variações genômicas podem compreender fatores atuantes na capacidade (PALATINI *et al.*, 2017; WHITFIELD *et al.*, 2017) e na competência vetorial (HANCOCK, PENELOPE A. *et al.*, 2018; SHAIKEVICH *et al.*, 2016).

Com o avanço da era genômica foi demonstrado que existem grandes porções não codificantes do DNA (sem potencial de gerar um produto protéico) nos mais variados organismos, que podem ser responsáveis por processos regulatórios (DAVIS *et al.*, 2018). Dentre os fatores para a resposta imune de artrópodes a regulação gênica, a nível de mRNA, envolve rotas de RNA de interferência (iRNA do inglês *Interference RNA*), sendo este mecanismo apontado como fator chave em resposta imune de culicídeos contra vírus (BLAIR; OLSON, 2015; BRONKHORST; VAN RIJ, 2014). Na última década, estudos têm demonstrado que parasitas genômicos, como transposons e Elementos Virais Endogenizados (EVEs) podem ser importantes agentes no sistema imune de culicídeos (PALATINI *et al.*, 2017; TASSETTO *et al.*, 2018; WHITFIELD *et al.*, 2017), devido ao seu potencial em servir como fonte de produção de iRNAs, como sRNAs (do Inglês *Small RNA*) e piRNAs (do Inglês *PIWI-interacting RNA*), demonstrando a complexa relação evolutiva entre culicídeos, vírus e EVEs.

2.2 Vírus e Arbovírus

Dois conceitos elaborados nos últimos anos têm sido utilizados para descrever os vírus quando uma classificação abrangente se faz necessária: os Organismos Codificadores de Ribossomos, fazendo referência a todos os organismos dos três domínios (Procariotos, Arqueas e Eucariotos), e os Organismos Codificadores de Capsídeos, fazendo referência à todos os vírus, bem como aos diversos parasitas genômicos, como transposons e EVEs, inseridos na subcategoria de *Capsidless Elements* (Organismos Codificadores de Capsídeos que perderam, ou não desenvolveram capsídeo) (KOONIN; DOLJA, 2014).

2.2.1 Taxonomia

Os vírus, então inseridos no grande grupo de Organismos Codificadores de Capsídeos, podem ser classificados de duas formas: Pelo tipo de material genético, sendo o modelo atualmente mais utilizado, tendo como base a classificação de Baltimore (BALTIMORE, 1970), ou pelo modelo do ciclo de expressão/replicação, sendo este último um novo modelo proposto por Dolja e Koonin (2014).

Neste trabalho foi utilizada a classificação de Baltimore, na qual consiste em sete classes virais (Figura 2):

- a) Classe I: Vírus de dupla fita de DNA (dsDNA, do Inglês *Double Strand DNA*);
- b) Classe II: Vírus de fita única de DNA com polaridade positiva, ou seja, com a mesma polaridade do mRNA viral (ssDNA, do Inglês *Single Strand DNA*);
- c) Classe III: Vírus de dupla fita de RNA (dsRNA, do Inglês *Double Strand RNA*);
- d) Classe IV: Vírus de fita única de RNA com polaridade positiva ((+)ssRNA, do Inglês *Single Strand RNA*, com o símbolo “+” indicando a polaridade da fita);
- e) Classe V: Vírus de fita única de RNA com polaridade negativa ((-)ssRNA, do Inglês *Single Strand RNA*, com o símbolo “-” indicando a polaridade da fita);
- f) Classe VI: Vírus de fita única de RNA com transcrição reversa (ssRNA-RT do Inglês *Single Strand RNA with Reverse Transcription*);
- g) Classe VII: Vírus de fita dupla de DNA com transcrição reversa (dsDNA-RT do Inglês *Double Strand DNA with Reverse Transcription*).

A classificação original de Baltimore descrevia apenas as classes de I a VI, porém em 1985 foram descritos vírus de DNA fita dupla com etapa de transcrição reversa (TEMIN, 1985), sendo adicionada uma sétima categoria à lista. Seguindo a classificação de Baltimore, o Comitê Internacional de Taxonomia Viral (ICTV do Inglês *International Committee on*

Taxonomy of Viruses) padronizou um sistema de classificação taxonômica com 8 níveis: Reino, Filo, Classe, Ordem, Família, Subfamília, Gênero e Espécie, onde diversas características podem ser utilizadas para classificação taxonômica, como identidade de nucleotídeos, compartilhamento de hospedeiro, estrutura genômica e outros aspectos da biologia viral.

Figura 2 - Sistema de classificação viral de Baltimore e sistema taxonômico proposto pelo ICTV

Baltimore	Principais Hospedeiros	ICTV**
Classe I dsDNA	Bactérias	Caudovirales
	Arquéias	Ligamenvirales
	Plantas, Vertebrados terrestres e aquáticos, invertebrados marinhos	*Herpesvirales
Classe II ssDNA	Plantas, Vertebrados e invertebrados terrestres	*Parvoviridae
Classe III dsRNA	Vertebrados e invertebrados terrestres	*Reoviridae
	Vertebrados e invertebrados terrestres	*Flaviviridae
Classe IV (+)ssRNA	Vertebrados e invertebrados terrestres, algas	*Picornavirales
	Vertebrados terrestres, invertebrados marinhos e terrestres	*Nidovirales
	Vertebrados e invertebrados terrestres	*Togaviridae
	Plantas	Tymovirales
Classe V (-)ssRNA	Plantas	Serpentovirales
	Invertebrados marinhos e terrestres	Muvirales
	Invertebrados terrestres e marinhos	*Jingchuvirales
	Vertebrados terrestres e aquáticos, plantas, invertebrados terrestres	*Mononegavirales
	Vertebrados e invertebrados terrestres	*Bunyavirales
	Vertebrados e invertebrados terrestres	*Articulovirales
Classe VI ssRNA-RT	Vertebrados e invertebrados terrestres e plantas	*Ortervirales
Classe VII ssDNA-RT		

Fonte: O autor, baseado em Yuri *et al.* (2017); Valles *et al.* (2017); Simmonds *et al.* (2017); Chen *et al.* (2018).
 Nota: * Representam grupos virais que já foram identificados infectando culicídeos, ** International Committee on Taxonomy of Viruses.

Além das classificações taxonômicas tradicionais, existem grupos de vírus englobados em categorias generalistas, como arbovírus e vírus específicos de insetos (ISVs, do Inglês *Insect Specific Viruses*). O termo Arbovírus não faz referência a um nível taxonômico específico, e sim a vírus que são transmitidos por artrópodes, como por exemplo: mosquitos, carrapatos e moscas. As famílias *Flaviviridae*, *Bunyaviridae* e *Togaviridae* englobam a maioria dos arbovírus, enquanto as famílias *Reoviridae* e *Orthomyxoviridae* possuem poucos representantes descritos nesta categoria (CLETON *et al.*, 2012). Além do hospedeiro invertebrado, que atua como vetor, hospedeiros vertebrados, como mamíferos e aves podem ser infectados com arbovírus. Dos mais de 530 Arbovírus descritos até 2016, pelo menos uma centena pode ser considerada patogênicos para humanos (LEQUIME; PAUL; LAMBRECHTS, 2016).

Caso o vírus possua a capacidade de infectar apenas insetos, e não outros grupos de vertebrados ou invertebrados, o vírus em questão é incluído no grupo de vírus específicos de insetos. ISV compreende uma nova categoria de vírus, considerada atualmente como parte da microbiota de insetos (ROUNDY *et al.*, 2017), em especial de mosquitos, nos quais os ISVs foram primeiramente (STOLLAR; THOMAS, 1975) e majoritariamente identificados (ROUNDY *et al.*, 2017).

2.2.2 Estrutura Genômica

De acordo com o ICTV, atualmente existem 14 ordens virais descritas (Figura 2), que apresentam características genômicas conservadas, conferindo, nos termos apontados em cada proposta de classificação, um grau de “parentesco” entre as famílias virais de determinada ordem, além de famílias virais que não estão classificadas em nenhuma ordem, por não possuírem características estruturais ou biológicas compartilhadas com outras famílias.

Com exceção das ordens *Caudavirales* e *Ligamenvirales*, que infectam bactérias e arqueas, respectivamente, todas as demais ordens possuem representantes que infectam eucariontes. Excluindo as ordens *Serpentovirales* e *Tymovirales*, com representantes que infectam apenas plantas, todas as demais possuem representantes que infectam animais vertebrados ou invertebrados. E no caso de representantes do grupo de arbovírus, as famílias *Flaviviridae*, *Reoviridae* e *Togaviridae* ainda não foram classificadas em nenhuma ordem (International Committee on Taxonomy of Viruses, ICTV, 2019).

2.2.2.1 Ordem Muvirales

Compreende apenas uma família viral, denominada *Qinviridae*, subdividida em um gênero e 8 espécies, com representantes que infectam tanto invertebrados marinhos quanto terrestres (SHI *et al.*, 2016).

Devido à recente definição desta ordem (YURI *et al.*, 2017), existem poucas informações relacionadas à estrutura genômica de seus membros, podendo ser definido apenas o seu tipo de material genético (RNA fita única de sentido negativo) que codifica Nucleoproteínas e RNA-polimerases dependentes de RNA (SHI *et al.*, 2016).

2.2.2.2 Ordem *Jingchuvirales*

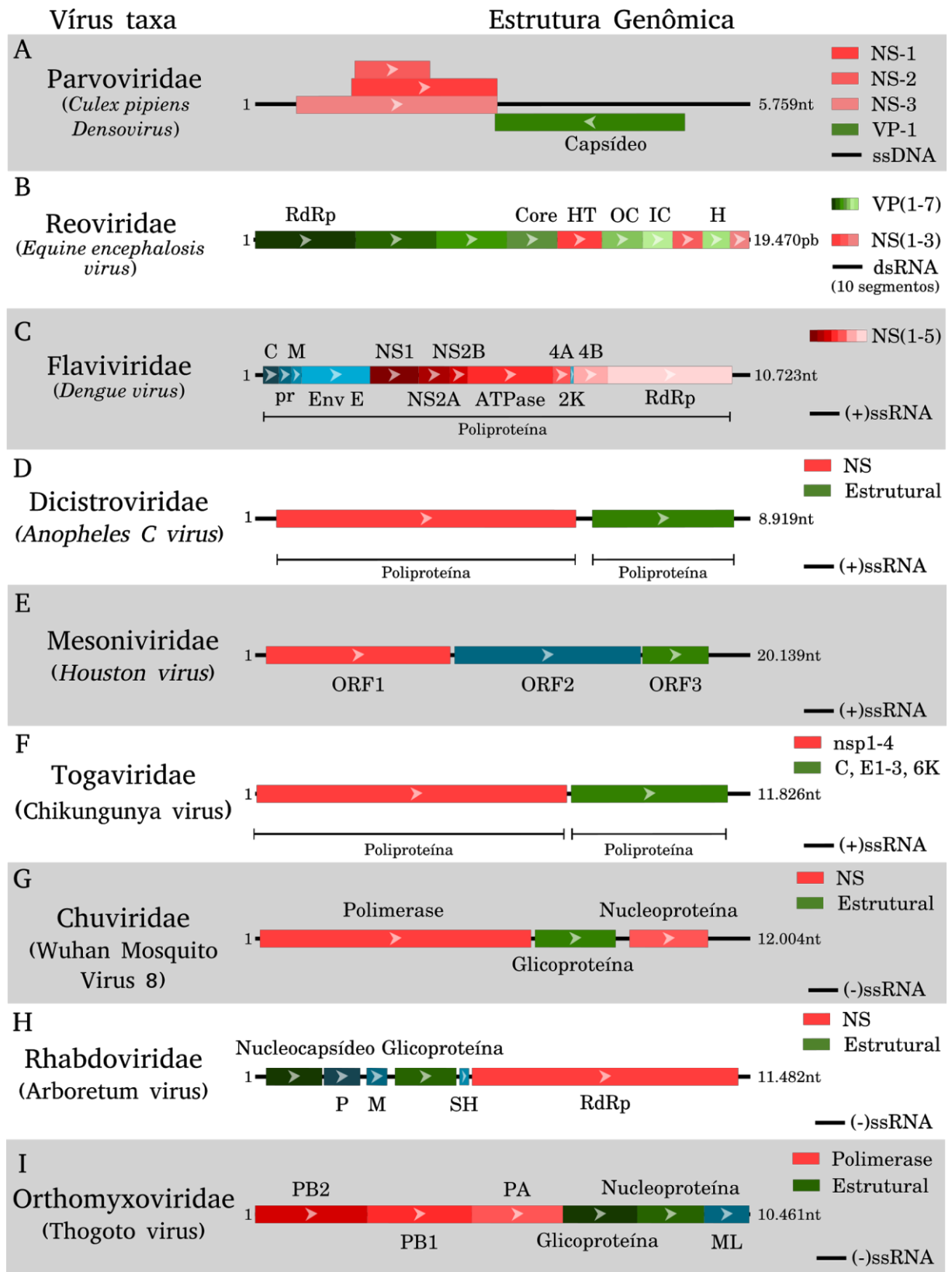
Relacionada à ordem *Mononegavirales*, é composta por apenas uma família (*Chuviridae*), um gênero e 29 espécies, com representantes que infectam invertebrados marinhos, como caranguejos por exemplo, e diversos insetos, incluindo mosquitos (LI, CI-XIU *et al.*, 2015).

Possuem como material genético uma molécula de RNA fita única sentido negativo, que pode estar estruturada de três formas: circular segmentado, circular não-segmentado ou linear, possuindo caracteristicamente ORFs que codificam Glicoproteínas, RNA-polimerases dependentes de RNA e, em alguns representantes, Nucleoproteínas (Figura 3) (LI, CI-XIU *et al.*, 2015)

2.2.2.3 Ordem *Mononegavirales*

Compreende ao todo 11 famílias virais: *Artoviridae*, *Bornaviridae*, *Filoviridae*, *Lispiviridae*, *Mymonaviridae*, *Nyamiviridae*, *Paramyxoviridae*, *Pneumoviridae*, *Rhabdoviridae*, *Sunviridae* e *Xinmoviridae*, subdivididas num total de 55 gêneros e 280 espécies. Possui representantes patogênicos à mamíferos, incluindo humanos, nas famílias *Bornaviridae* (TIZARD *et al.*, 2016), *Filoviridae* (família que engloba o gênero *Ebolavirus*) (BEER; KURTH; BUKREYEV, 1999), *Paramyxoviridae* (PHILBEY *et al.*, 1998) e *Rhabdoviridae* (BASAK *et al.*, 2007). Os vírus da família *Rhabdoviridae* (Figura 3) são distribuídos em diversos grupos de hospedeiros, desde plantas e vertebrados aquáticos e terrestres até invertebrados, incluindo mosquitos (COFFEY *et al.*, 2014; GULERIA *et al.*, 2011).

Figura 3 - Representação esquemática da estrutura genômica de alguns representantes virais já relatados infectando mosquitos. Com base em mapas genômicos presentes no NCBI (Apêndice A)



Fonte: O autor.

Vírus dessa ordem possuem genoma linear de RNA fita única senso negativo (Classe V) tendo um tamanho médio entre 8,9 e 19 kb e codificando de 5 a 7 proteínas estruturais, como por exemplo glicoproteínas, proteínas de matriz de membrana, proteínas de ligação ao RNA. Além do conteúdo codificante, outras características como complementaridade inversa das regiões 3' e 5' e expressão individual dos genes servem como fatores para classificar vírus desta ordem (PRINGLE, 1996).

2.2.2.4 Ordem *Bunyavirales*

Doze famílias virais: *Arenaviridae*, *Cruliviridae*; *Fimoviridae*; *Hantaviridae*; *Leishbuviridae*; *Myoviridae*; *Nairoviridae*; *Peribunyaviridae*; *Phasmaviridae*; *Phenuiviridae*; *Tospoviridae* e *Wupedeviridae* compoem, atualmente, a ordem *Bunyavirales*, contando com 46 gêneros e 291 espécies (MAES *et al.*, 2018). Destas doze famílias, pelo menos quatro possuem representantes apontados como patogênicos à humanos (*Nairoviridae*, *Phenuiviridae*; *Peribunyaviridae* e *Hantaviridae*), sendo a família *Hantaviridae* a única não englobada na categoria de Arbovírus (GUARDADO-CALVO; REY, 2017).

Vírus desta ordem apresentam como material genético uma fita única de RNA com polaridade negativa (Classe V), variando de 12 a 19 kb, segmentado em três regiões: S, M e L (do Inglês *Small*, *Middle* e *Large*), sendo os segmentos S e L responsáveis por codificar as proteínas (Nucleoproteína e RNA-Polimerase dependente de RNA) envolvidas na replicação do genoma viral, e o segmento M responsável por codificar as Glicoproteínas (Gn e Gc) do capsídeo viral (GUARDADO-CALVO; REY, 2017).

2.2.2.5 Ordem *Articulavirales*

Compreende duas famílias, 8 gêneros e 10 espécies, com representantes que infectam tanto animais vertebrados, no caso de vírus da família *Amnoonviridae* (EYNGOR *et al.*, 2014), quanto invertebrados no caso dos representantes do gênero *Thogotovirus* da família *Orthomyxoviridae*, que são considerados arbovírus (HAIG; WOODALL; DANSKIN, 1965).

Vírus desta ordem apresentam genoma de RNA fita única senso negativo, com número variado de segmentações, variando de 8 segmentos, no caso dos vírus Influenza a 6 segmentos no caso dos vírus Thogoto, com genomas variando de 10 à 14 kb. Mesmo com a divergência de hospedeiros, algumas proteínas são comuns a todos os gêneros desta ordem,

como RNA-polimerase dependente de RNA, Nucleoproteína, Hemaglutinina, Glicoproteína e proteínas matrizes não glicosiladas (KNIPE; HOWLEY, 2013; (PRESTI *et al.*, 2009).

2.2.2.6 Ordem *Herpesvirales*

Subdividida em três famílias virais: *Alloherpesviridae*, *Herpesviridae* e *Malacoherpesviridae*, a ordem *Herpesvirales* engloba, atualmente, 19 gêneros e 149 espécies (DAVISON *et al.*, 2009). Os membros da família *Herpesviridae* possuem representantes patogênicos à aves, répteis e mamíferos, incluindo humanos, enquanto peixes e anfíbios podem ser infectados por vírus da família *Alloherpesviridae* e bivalves por *Malacoherpesviridae*. Dessa forma, não existem relatos de vírus da ordem *Herpesvirales* infectando insetos.

Representantes dessa ordem possuem um genoma linear de DNA dupla fita, fazendo parte da Classe I juntamente com a ordem *Caudovirales*, variando de 125 à 295 kbp. Existem diferentes organizações de genomas tanto na questão de regiões repetitivas, quanto na ordem e presença de diferentes genes, o que resulta na produção entre 70 e mais de 200 proteínas dependendo do vírus (KNIPE; HOWLEY, 2013).

2.2.2.7 Ordem *Nidovirales*

Compreende 9 famílias virais, 33 gêneros e 88 espécies (GORBALENYA *et al.*, 2017, ICTV 2019). Vírus dessa ordem possuem uma gama diversificada de hospedeiros, desde vertebrados como mamíferos e aves no caso de vírus da família *Coronaviridae*, quanto invertebrados tanto marinhos, como vírus da família *Abyssoviridae* (BUKHARI *et al.*, 2018), quanto terrestres, no caso da família *Mesoniviridae*, sendo esta última família viral identificada em análises metagenômicas infectando mosquitos da espécie *Culex quinquefasciatus* (CHARLES *et al.*, 2018).

Vírus dessa ordem possuem um genoma de RNA fita única sentido positivo e podem ser divididos em dois grupos, os Nidovírus de genoma grande, variando de 26 a 31 kb, e Nidovírus de genoma pequeno, variando de 12 a 17 kb. Apesar das diferenças de tamanho do genoma, e da diversidade de hospedeiros, estes vírus apresentam um conjunto de proteínas conservadas: Glicoproteínas (*Spyke*, *Large*, *Major*, *Minor* e *Small*), de Membrana, Nucleocapsídeo, Envelope e Hemaglutinina-esterase (GORBALENYA *et al.*, 2006; PASTERNAK, 2006).

2.2.2.8 Ordem *Ortervirales*

Engloba todos os vírus que apresentam uma etapa de transcrição reversa, sendo subdividida em 5 famílias: *Retroviridae*, *Caulimoviridae*, *Belpaoviridae*, *Metaviridae*, *Pseudoviridae*, sendo as três últimas famílias também consideradas superfamílias de elementos transponíveis (BEL, Gypsy e Cópia, respectivamente) (KAPITONOV; JURKA, 2007). No que diz respeito à diversidade de hospedeiros, estes vírus já foram identificados infectando plantas, no caso de vírus da família *Caulimoviridae*, animais vertebrados, desde aves até primatas humanos e não-humanos, como vírus da família *Retroviridae*, e invertebrados, incluindo mosquitos, como vírus das famílias *Belpaoviridae* e *Pseudoviridae* (KINSLEY *et al.*, 2016; KRUPOVIC *et al.*, 2018; TU; HILL, 1999).

Ao contrário de outras ordens virais, esta ordem não agrupa os vírus pelo tipo de material genético, mas sim pelo mecanismo de retrotranscrição presente no ciclo de replicação destes vírus. São identificados então, dois tipos de material genético nesta ordem, vírus com genoma de dupla fita de DNA, no caso das famílias *Caulimoviridae* e *Retroviridae* (subfamília *Spumaretrovirinae*), e de RNA fita única com sentido positivo no caso das famílias *Belpaoviridae*, *Metaviridae*, *Pseudoviridae* e *Retroviridae*. Apesar da divergência no tipo de material genético, existe uma conservação de estruturas entre os vírus da ordem *Ortervirales*, como por exemplo, a presença de uma poliproteína com informações para gerar uma Protease, uma Transcriptase Reversa, uma Integrase e uma RNase H (KRUPOVIC *et al.*, 2018).

2.2.2.9 Ordem *Picornavirales*

Possui representantes em 6 famílias: *Dicistroviridae*, *Iflaviridae*, *Marnaviridae*, *Picornaviridae*, *Polycipiviridae* e *Secoviridae*, totalizando 64 gêneros e 260 espécies. Com representantes capazes de infectar tanto vertebrados como aves e mamíferos, incluindo humanos (vírus da Hepatite A) da família *Picornaviridae*, quanto algas (família *Marnaviridae*) e artrópodes, como abelhas e mosquitos, no caso de vírus da família *Dicistroviridae* (CARISSIMO *et al.*, 2016; VALLES *et al.*, 2017).

Vírus desta ordem apresentam como material genético uma fita única ou bipartida de RNA sentido positivo, variando de 7 a 12 kb com poliadenilação na extremidade 3' (EHRENFELD; DOMINGO; ROOS, 2010). Com informações para codificar proteínas de

capsídeo viral e três proteínas envolvidas na replicação destes vírus: Helicase, Protease de Cisteína e RNA-polimerase dependente de RNA (VALLES *et al.*, 2017).

2.2.2.10 Família *Flaviviridae*

Subdivida em quatro gêneros: *Flavivirus*, *Hepacivirus*, *Pegivirus* e *Pestivirus* e 89 espécies, esta família abriga arbovírus de grande importância para saúde humana, como por exemplo os vírus causadores da Dengue, Zika e Febre Amarela, bem como outros vírus de importância médica, como os Hepacivirus, causadores da Hepatite C e de doenças suínas, com no caso dos Pestivirus (GRARD *et al.*, 2010; KNIPE; HOWLEY, 2013).

Vírus dessa família possuem um genoma de RNA fita única com senso positivo, variando de 8,9 a 13 kb dependendo do gênero, com ausência de uma cauda poli-A na extremidade 3', e com a presença de uma estrutura de Cap na extremidade 5' no caso dos vírus do gênero *Flavivirus*. Apresentam em sua estrutura informações para codificar uma poliproteína, que após ser clivada, dá origem a proteínas estruturais (proteínas de capsídeo e envelope) e não-estruturais (Protease de Serina, Helicase e RNA-polimerase dependente de RNA) (SIMMONDS *et al.*, 2017; KNIPE; HOWLEY, 2013).

2.2.2.11 Família *Reoviridae*

Compreende 15 gêneros e 91 espécies, com um amplo espectro de hospedeiros, como por exemplo os vírus dos gêneros *Orthoreovirus* e *Rotavirus* que infectam vertebrados, os arbovírus dos gêneros *Orbivirus*, *Coltivirus* e *Seadornaviruses* já identificados em mosquitos e carrapatos, vírus do gênero *Cypovirus* que infectam insetos e do gênero *Phytoreovirus*, *Fijivirus* e *Oryzavirus* capazes de infectar plantas (NIBERT; FURLONG; FIELDS, 1991; KNIPE; HOWLEY, 2013).

Possuem genoma de RNA dupla fita, variando de 9 a 12 segmentos, com tamanho médio variando de 19 kbp no caso de vírus do gênero *Orbivirus* até 33 kbp em vírus do gênero *Cypovirus*. Dentre as estruturas conservadas entre os gêneros, pode-se citar os genes que codificam RNA-polimerase dependente de RNA, Fosfohidrolases, Guanylyltransferase, Transmethylases, Helicase e Pyrophosphatase. Além das porções codificantes, as fitas de RNA sentido positivo apresentam uma região de capeamento na extremidade 5' e ausência de cauda poli-A nas extremidades 3' (KNIPE; HOWLEY, 2013).

2.2.2.12 Família *Togaviridae*

Abrange 1 gênero (*Alphavirus*) e 31 espécies. No caso dos vírus do gênero *Alphavirus*, pode-se citar alguns arbovírus de importância média como o vírus Chikungunya, envolvido em epidemias globais, e os vírus Mayaro, Onyong-nyong, Ross River, e os vírus de encefalite equina (Venezuelano, do Leste e do Oeste) já reportados infectando tanto humanos quanto mosquitos (CHEN, RUBING *et al.*, 2018).

Apresentam como material genético uma fita única de RNA sentido positivo, variando de 9 a 11,8 kb, contendo conservadoramente um capeamento na região 5' e uma cauda poli-A na região 3', e informações para codificar 4 proteínas não estruturais, denominadas nsP1 até nsP4, que são necessárias para estruturação da maquinaria de replicação viral, e proteínas de capsídeo e glicoproteínas, necessárias para estruturação das partículas virais maduras (CHEN, RUBING *et al.*, 2018; KNIPE; HOWLEY, 2013).

2.2.2.13 Família *Parvoviridae*

Compreende duas subfamílias, *Densovirinae* e *Parvovirinae*, 13 gêneros e 80 espécies virais. Vírus da subfamília *Parvovirinae* possuem a capacidade de infectar vertebrados terrestres, como aves e mamíferos, incluindo humanos, enquanto vírus da subfamília *Densovirinae* possuem como hospedeiro diversos grupos de insetos, como lepidópteros e dípteros, onde vírus do gênero *Densovirus* já foram reportados infectando mosquitos dos gêneros *Anopheles* e *Culex* (BAQUERIZO-AUDIOT *et al.*, 2009; REN; HOICZYK; RASGON, 2008).

Possuem como material genético fita única de DNA não segmentado, variando entre 4 e 6,3 kb, apresentando regiões palindrômicas nas regiões 5' e 3' com tamanho variando entre 120 e 550 nucleotídeos. Ambas as subfamílias apresentam duas estruturas gênicas (REP e CP), que são lidas em duas ORFs no caso de vírus da subfamília *Parvovirinae*, e 4 ORFs no caso de vírus da subfamília *Densovirinae*. A estrutura REP codifica proteínas relacionada à replicação viral, enquanto a estrutura CP codifica proteínas relacionada à estrutura do capsídeo viral, os grupos e números de proteínas variam de acordo com o gene viral (KNIPE; HOWLEY, 2013).

Apesar das similaridades entre diversos grupos virais, que permitem estabelecer relações evolutivas e realizar classificações taxonômicas a nível de gênero, não foram identificados, até o momento, genes que estejam presentes em todos os grupos virais, dessa

forma, diferente dos organismos codificadores de ribossomos, os organismos codificadores de capsídeo possivelmente apresentam uma origem polifilética, tendo como representação evolutiva uma teia, com compartilhamento de alguns genes conservados entre alguns grupos, conhecidos como genes *hallmarks* (KOONIN; DOLJA, 2014; KOONIN; DOLJA; KRUPOVIC, 2015).

2.2.3 Relações Evolutivas

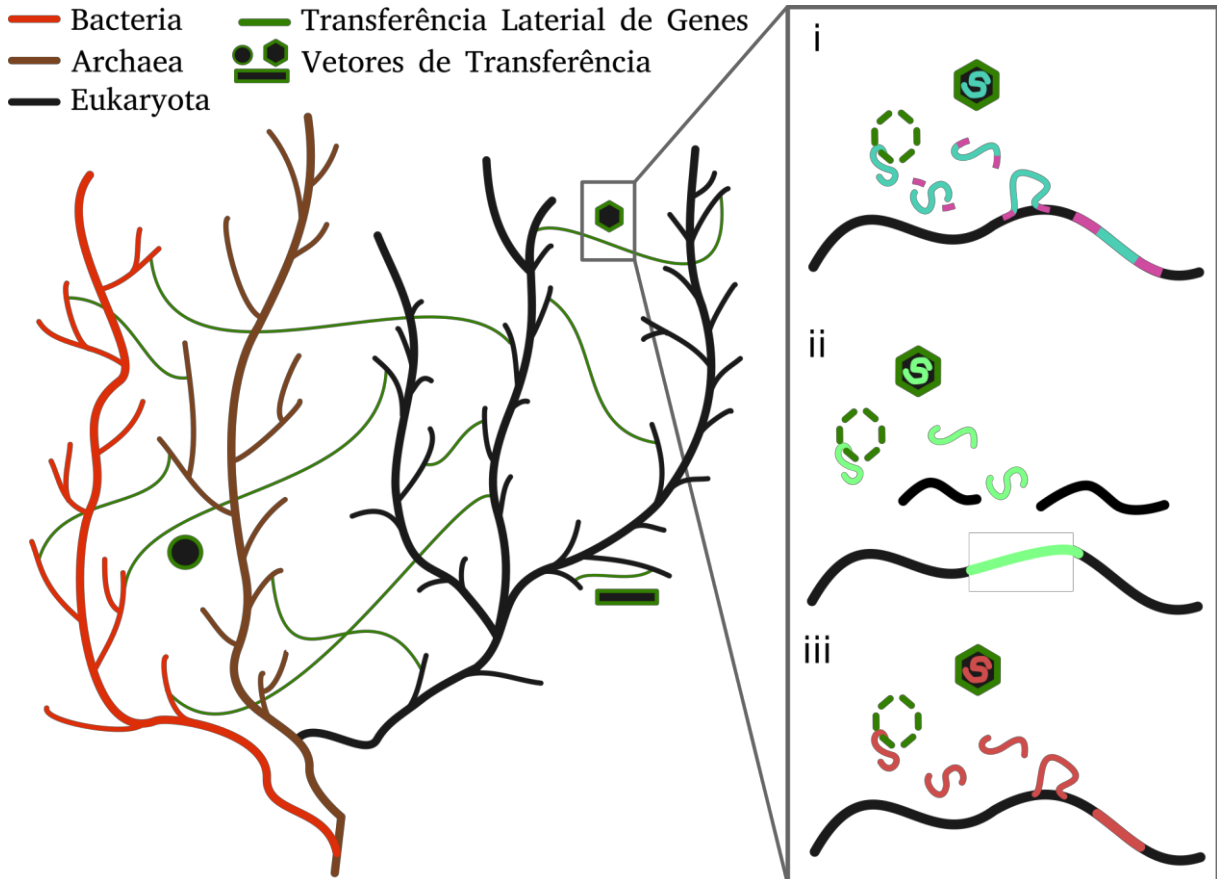
Na metade do século XIX, Charles Darwin e Alfred Wallace publicaram obras com bases na ideia do surgimento de novas espécies por meio de seleção natural, a qual seria regida por dois grandes fatores: a pressão que o meio exerce sobre o organismo; e a herança parental de indivíduos adultos bem sucedidos, para sua prole (WALLACE, 1855; DARWIN, 1859). Cinco anos após as publicações de Darwin e Wallace, Gregor Mendel publicou seus primeiros resultados demonstrando o papel de fatores, do que conhecemos hoje como genes, na herança de caracteres fenotípicos (MENDEL, 1864).

Até metade do século XX, aceitava-se a ideia que os genes eram transmitidos apenas verticalmente, ou seja, de indivíduos de uma geração, digamos F₀, para sua prole na geração F₁, até que, em 1947, surgiram os primeiros trabalhos demonstrando transferências laterais de genes entre bactérias, num processo que veio a ser descrito como Transferência Horizontal de Genes, ou seja, um processo de transferência gênica sem relação de parentesco, em primeiro grau, entre os indivíduos (TATUM; LEDERBERG, 1947).

No final do século XX, surgiram estudos evidenciando processos de transferência horizontal entre eucariotos, demonstrando que a história genômica evolutiva não compreende uma árvore linear, onde os genomas evoluem de forma vertical, e sim de uma árvore com várias teias, interligando diversos ramos por meio de transferências horizontais de genes (Figura 4) (BAHN, 2018).

Sabe-se que em bactérias a fonte para transferência horizontal advém de três categorias de vetores: plasmídeos, bacteriófagos e transposons. Em eucariotos tem-se um panorama similar, onde plasmídeos podem ser a fonte de transferência gênica, no caso de plantas e leveduras; transposons em praticamente todos os níveis taxonômicos (SOUCY; HUANG; GOGARTEN, 2015) e nos últimos anos têm-se verificado o papel de vírus nestes processos de transferência gênica. Até o presente estudo, são conhecidos mais de 1800 eventos de transferência horizontal de genes mediados por vírus em eucariotos (DOTTO *et al.*, 2018).

Figura 4 - Representação do panorama evolutivo entre os organismos. Enfatizando os três mecanismos de endogenização viral



Fonte: O autor, baseado em Katzourakis e Gifford (2010).

Legenda: (i) Integração por interação com transposons ativos; (ii) Integração por recombinação não-homóloga; (iii) Integração retroviral.

2.3 Elementos virais endogenizados

Eventos de transferência gênica englobam não só a transferência do gene em questão, mas também de regiões específicas ou da sequência genômica completa do vetor de transferência. Caso o vetor seja um vírus, o evento envolvendo a integração de alguma região genômica do vetor é classificado como endogenização viral, e o processo pode ocorrer por três principais vias: no caso de Retrovírus (Figura 4, item iii), o próprio ciclo de replicação viral propicia o evento de endogenização, visto que a integração do genoma do hospedeiro é essencial para a replicação deste grupo viral, uma falha na excisão, devido a mutações do vírus ou por meio de mecanismos de defesa epigenéticos, pode ocorrer, e o retrovírus em questão passa então a fazer parte do genoma, tornando-se um Retrovírus Endógeno (ERV) (PARSEVAL; HEIDMANN, 2005).

Nos demais grupos virais, que não apresentam etapa de integração viral obrigatória durante seu ciclo replicativo, a endogenização viral pode ocorrer de duas formas: através de eventos de recombinação não-homóloga (Figura 4, item ii), ou por integração mediada por maquinaria de transposons ativos (Figura 4, item i). Vírus ou fragmentos virais que sofrem este processo de integração no genoma hospedeiro são conhecidos como Elementos Virais Endogenizados. Caso estes eventos ocorram nas células germinativas, os EVEs passam a fazer parte do genoma da prole do indivíduo e são herdados verticalmente para todos os descendentes do hospedeiro em questão (KATZOURAKIS; GIFFORD, 2010).

A integração de um fragmento viral pode ser vista de três diferentes formas quando levamos em consideração o *fitness* (desempenho do organismo frente à pressões do ambiente ou competição com outros membros da população) do hospedeiro: I - Neutro, II - Seletivamente vantajoso ou III - Desvantajoso. Caso a integração viral resulte em uma característica vantajosa ao hospedeiro, pode ocorrer um processo de domesticação genética, onde o gene de origem exógena torna-se um gene essencial para o sucesso da espécie aceptora, podendo manter a mesma função ou ser selecionado para uma nova função, num processo conhecido como exaptação, proposto por Gould em 1981 para caracteres estruturais (GOULD; VRBA, 1981) e aplicado à nível molecular nas últimas décadas (ALLEN *et al.*, 2004; JOSE; GARDNER; BARQUIST, 2019).

EVEs podem causar uma variedade de impactos no genoma hospedeiro como: I) gerar novas mutações devido a inserção e reinserção do genoma viral resultantes de vírus endogenizados que ainda possuem a capacidade de se replicar dentro do genoma hospedeiro (ARMEZZANI *et al.*, 2014); II) o aumento ou diminuição da expressão de genes do hospedeiro devido à extensão de marcas epigenéticas que controlam a expressão e replicação viral e que acabam por controlar a expressão gênica de genes flanqueadores (LAMPRECHT *et al.*, 2010); III) mudar o padrão de recrutamento de fatores de transcrição pelos promotores virais que podem modificar a expressão gênica de genes flanqueadores (FRANK; FESCHOTTE, 2017).

Alguns exemplos desses efeitos podem ser encontrados em retrovírus endógenos de humanos. Esses vírus são sistematicamente silenciados por marcadores epigenéticos tanto em células embrionárias como somáticas, mas devido a falhas para manter esse controle ocorre a reativação dos ERVs a nível transcricional, causando a ativação de um proto-oncogene flanqueador responsável pelo desenvolvimento do Linfoma de Hodgkin's (LAMPRECHT *et al.*, 2010).

Apesar da possibilidade de gerar consequências desastrosas para o genoma, através dos mecanismos reportados acima, os produtos gênicos dos remanescentes virais podem ser úteis ao organismo hospedeiro. Alguns EVEs são conhecidos por produzir proteínas que oferecem imunidade contra novas infecções virais em nível pós traducional. Por exemplo, o vírus caprino Jaagsiekte endogenizado (enJSRV - *endogenous Jaagsiekte sheep retrovírus*) é capaz de proteger o trato genital de ovelhas em pelo menos dois níveis: I) produzindo uma proteína do envelope viral que se liga aos receptores celulares usados pelas partículas virais do JSRV para entrar nas células epiteliais do trato genital, fazendo com que ocorra uma diminuição substancial na entrada de novas partículas virais; II) enJSRV produz uma proteína Gag que tem um dobramento da sua estrutura tridimensional errôneo, apesar disso, essa proteína é capaz de ser utilizada na produção de novas partículas virais que serão alvos de degradação dos proteossomos devido à sua natureza quimérica (proteínas Gag dobradas corretamente e mal dobradas) (ARMEZZANI *et al.*, 2014).

Outro mecanismo antiviral pode ocorrer a nível pós-transcricional, onde vírus endógenos atuam como fonte na produção de iRNAs como sRNAs e piRNAs, e estes iRNAs atuam ativamente na resposta antiviral. Este mecanismo de defesa antiviral tem sido elucidado em mosca da fruta (*Drosophila melanogaster*) (MARQUES *et al.*, 2013) e alguns trabalhos vêm investigando o potencial destes EVEs em sistema imune de culicídeos (WHITFIELD *et al.*, 2017).

Alguns ERVs de mamíferos têm a capacidade de reforçar o sistema imunológico de seu hospedeiro através de suas regiões longas repetidas (LTR, do Inglês *Long Terminal Repeat*), que contém *enhancers* capazes de induzir a produção de *interferons*, podendo contribuir em alguns casos para regular genes adjacentes do hospedeiro que codificam fatores críticos ao sistema imune (FRANK; FESCHOTTE, 2017). E de atuar no desenvolvimento da placenta no grupo de mamíferos da infraclasse *Eutheria*, onde a endogenização e subsequente exaptação de um capsídeo retroviral deu origem aos genes *sync*, responsáveis pela formação de diversos tecidos constituintes da placenta (BLAISE *et al.*, 2003; DUPRESSOIR *et al.*, 2005; MI *et al.*, 2000).

O primeiro relato de EVEs em culicídeos ocorreu devido à identificação de genes virais em diversas culturas de célula C6/36 de *Aedes albopictus*, onde um segmento genômico apresentou alta similaridade com genes (NS1, NS4A e NS4B) de vírus da família *Flaviviridae* (*Cell Fusing Agent* e *Kamiti River*) (CROCHU *et al.*, 2004). Desde então EVEs da família *Rhabdoviridae* foram identificados em genoma de *Aedes aegypti* e *Culex quinquefasciatus* com estruturas intactas de ORF (do Inglês *Open Reading Frames*) de

Nucleoproteínas e Glicoproteínas (KATZOURAKIS; GIFFORD, 2010), outros EVEs da família *Flaviviridae* foram identificados em *Anopheles minimus* contendo diferentes regiões gênicas (LEQUIME; LAMBRECHTS, 2017) e em diferentes populações de *Aedes aegypti*, onde não foi possível detectar proteínas derivadas destes EVEs, porém foram identificados piRNAs que apresentaram relação com estes EVEs (PALATINI *et al.*, 2017; WHITFIELD *et al.*, 2017).

Em uma meta-análise de genômica comparativa foram identificados 242 EVEs derivados de vírus das famílias *Reoviridae*, *Flaviviridae*, *Bunyaviridae* e *Rhabdoviridae* num total de 22 espécies de culicídeos, onde os EVEs de *Aedes aegypti* das famílias *Flaviviridae* e *Rhabdoviridae* apresentaram, por análises de bioinformática (PALATINI *et al.*, 2017) e por experimentos *in vitro* (TASSETTO *et al.*, 2018; WHITFIELD *et al.*, 2017), a capacidade de produzir piRNAs, demonstrando o potencial destes parasitas genômicos em atuar no sistema de resposta antiviral em culicídeos.

Apesar do estudo destes elementos genéticos em culicídeos, o foco, até o presente estudo, foi avaliar as atividades destas sequências em processos de regulação gênica por mecanismos de pequenos RNAs não codificantes. Dessa forma, a caracterização a nível estrutural desses elementos virais endogenizados, bem como sua capacidade codificante e as proteínas produzidas constituem um recurso valioso tanto para a reconstrução da história evolutiva dos vírus, quanto para entender as possíveis implicações destes elementos nos genomas de seus hospedeiros (FESCHOTTE; GILBERT, 2012).

2.4 Paleovirologia

Quando um vírus sofre um processo de endogenização e subsequentemente passa a fazer parte do genoma do hospedeiro, novas forças evolutivas passaram atuar sobre o EVE. Diversos fatores atuam sobre a taxa mutacional de um vírus, como por exemplo: A polimerase envolvida na replicação, o tamanho do genoma viral, pressões que a célula hospedeira impõem sobre o patógeno invasor e até mecanismos da própria estratégia de replicação viral (DUFFY; SHACKELTON; HOLMES, 2007). Quando o vírus passa a fazer parte do genoma hospedeiro, estes fatores não mais atuam sobre a taxa de mutação viral e o genoma/fragmento viral passa a evoluir sob influência da variabilidade e pressão seletiva do genoma hospedeiro (EMERMAN; MALIK, 2010).

Em eucariotos, a pressão evolutiva dependerá de diversos fatores: região em que o vírus foi inserido, podendo ser uma região conservada (com baixas taxas mutacionais) ou

uma região variável (com altas taxas mutacionais); a nova função viral dentro do genoma hospedeiro também é um fator crucial às pressões evolutivas, podendo o EVE, em um primeiro momento, sofrer pressões positivas (onde a seleção natural favorece mutações que resultam em substituição do aminoácido correspondente ao códon), pressões negativas (onde mutações sinônimas, que não resultam em substituição de aminoácido, são favorecidas); e mecanismos de proteção do próprio genoma eucarioto podem atuar sobre o EVE, como mecanismos epigenéticos, por exemplo (DUFFY; SHACKELTON; HOLMES, 2007) ou pressões neutras, quando a inserção não ocasiona nenhum efeito significativo no *fitness* do hospedeiro, ocasionando um processo conhecido como evolução neutra (KIMURA; OHTA, 1974).

Pressões positivas estão relacionadas à dois possíveis eventos: ao processo de exaptação, ou seja, quando o gene viral endogenizado está sob seleção diversificadora atuando uma nova função dentro do genoma hospedeiro em que a variabilidade a nível de aminoácido traz vantagens evolutiva ao hospedeiro (JOLY-LOPEZ; BUREAU, 2018), ou simplesmente o EVE em questão foi inserido numa região variável do genoma e não possui nenhum papel que contribuiu para o *fitness* do hospedeiro (evolução neutra). O mesmo pode ocorrer com eventos de pressões negativas, onde o EVE pode ter passado por um processo de domesticação, e pressões negativas atuam para manter a função do gene viral, como ocorre com genes *housekeeping*, ou o EVE pode simplesmente ter sido inserido numa região com baixa taxa mutacional. A relevância evolutiva do EVE só pode então ser mensurada com estudos evolutivos (identificação do EVE em diversos organismos) e estudos funcionais, avaliando a estrutura das proteínas derivadas do EVE, a ontologia dos genes e as possíveis interações em redes gênicas e regulatórias.

Independente da pressão exercida pelo genoma eucarioto, a taxa mutacional será, em média, menor que a de vírus circulantes, podendo chegar em um número de até um milhão de vezes menor de substituição de nucleotídeos por ano, quando comparamos a taxa mutacional média de vírus de RNA que se encontra entre 1×10^{-2} /base/ano a 1×10^{-5} /base/ano (DUFFY; SHACKELTON; HOLMES, 2007) com a taxa mutacional de genomas eucariotos que é em média $0,5 \times 10^{-9}$ /pb/ano (SCALLY, 2016). Devido a esta diferença entre taxas mutacionais, EVES vêm sendo considerados como vírus fósseis ou paleovírus, por manterem estruturas nucleotídicas correspondentes a vírus ancestrais e área de estudo que tem crescido no campo de análise de paleovírus veio a ser denominada de Paleovirologia (ASWAD; KATZOURAKIS, 2015; EMERMAN; MALIK, 2010; KATZOURAKIS, 2017; PATEL; EMERMAN; MALIK, 2009; TSANGARAS; GREENWOOD, 2017).

2.5 Importância de EVEs em análises de Metavirômica

Com o barateamento das tecnologias de sequenciamento, novas abordagens para detecção viral em larga escala foram desenvolvidas, culminando no desenvolvimento de uma área denominada metavirômica (KOONIN; DOLJA, 2018). Através do sequenciamento em larga escala do RNA total dos organismos seguido da busca por assinatura virais, diversos estudos aumentaram nosso conhecimento sobre diversidade viral de forma exponencial nos últimos cinco anos (LI, CI-XIU *et al.*, 2015; SHI *et al.*, 2016, 2018), especialmente em culicídeos (SHI *et al.*, 2017; XIA *et al.*, 2018).

Dentre os vírus que infectam mosquitos, os grupos que possuem genoma de RNA (ssRNA(-), ssRNA(+) e dsRNA) compoem a grande maioria de vírus identificados nestes organismos (NOURI *et al.*, 2018). Dessa forma, o sequenciamento de RNA (RNA-seq), é uma das principais abordagens utilizadas em análises de metavirômica (NOURI *et al.*, 2018). Entretanto, a detecção de RNA viral em dados de RNA-seq não garante que o vírus detectado seja um vírus capaz de se replicar e infectar outras células. O RNA viral detectado pode ser proveniente de fragmentos virais endogenizados no genoma e que são ativamente transcritos. Dessa forma, o conhecimento do repertório de vírus endogenizados um ponto essencial em estudos de metavirômica pois permite discernir entre vírus endogenizados e vírus infectivos (NOURI *et al.*, 2018).

2.6 Bioinformática como ferramenta de análise de vírus endogenizados

Assim como na Paleontologia são necessárias uma série de ferramentas para extrair o fóssil da rocha, na Paleovirologia existe um conjunto de ferramentas que permitem ao pesquisador localizar, identificar e extrair a informação de vírus endogenizados, permitindo separar com precisão o material genético dos EVEs e de seus hospedeiros. Estas ferramentas tratam-se de *softwares* e *scripts* que vêm sendo desenvolvidos nas últimas duas décadas, frutos da “Era Genômica”, e classificadas atualmente como ferramentas de bioinformática, como por exemplo a ferramenta BLAST (ALTSCHUL, STEPHEN F *et al.*, 1989), que busca por similaridade entre as sequências, HMMER (FINN; CLEMENTS; EDDY, 2011) que utiliza modelos ocultos de Markov para detectar padrões estruturais de sequências, e ferramentas que buscam padrões de características de elementos, onde não existem, até o momento, ferramentas para detecção direta de vírus endógenos, mas existem ferramentas para detecção de *transposons*, que permitem, em alguns casos, recuperar retrovírus

endógenos, como a ferramenta RepeatExplorer (NOVÁK *et al.*, 2012) e DNApipeTE (GOUBERT *et al.*, 2015), por exemplo.

Estudos têm demonstrado que comparações à nível de aminoácido apresentam sensibilidade superior, entre 30 e 100 vezes, quando comparado à comparações a nível de nucleotídeos (ABASCAL; ZARDOYA; TELFORD, 2010), dessa forma, estratégias de busca de similaridade utilizando regiões protéicas tornam-se mais indicadas quando se pretende identificar e comparar sequências altamente divergentes, como vírus circulantes e EVEs.

2.6.1 Análises filogenéticas: reconstruindo a história evolutiva de sequências genômicas

Os melhores métodos para definir taxonomicamente a origem de uma sequência são os métodos filogenéticos, uma vez que partindo de um alinhamento, são utilizados diferentes tipos de estratégias, levando em conta um modelo evolutivo para reconstruir a história evolutiva do objeto de estudo, ou seja, trata-se de avaliar a similaridade de sequências homólogas, partindo, então, do pressuposto que as sequências que estão sendo analisadas, sejam de DNA ou de proteínas, provém de um ancestral comum (BALDAUF, 2002). Isso permite não só a comparação das sequências com base em similaridade, mas também definir possíveis ancestrais do objeto de estudo (AJAWATANAWONG, 2016)

Existem 5 principais métodos de inferência evolutiva: UPGMA, Aproximação de vizinhos (NJ, do Inglês *Neighbor Joining*), Máxima Parcimônia, Máxima Verossimilhança e Estatística Bayesiana (VERLI, 2014). Estes métodos podem ser subdivididos em duas categorias: Métodos quantitativos, ou de distância, no caso de UPGMA e NJ, e métodos qualitativos, que engloba os demais métodos. Os métodos quantitativos são baseados em cálculos de matriz de distância em um conjunto de sequências alinhadas, desconsiderando modelos evolutivos e conseqüentemente formando árvores que não condizem com a história evolutiva das sequências.

Métodos qualitativos aplicam modelos evolutivos em um conjunto de sequências alinhadas e buscam apresentar uma árvore condizente com a história evolutiva do conjunto de sequências estudadas. Apesar da aplicação de modelos evolutivos, o método de parcimônia pode não representar a história evolutiva real, visto que busca criar uma árvore com o menor número de “passos evolutivos”, baseado num princípio que a evolução opta por caminhos mais simples, teoria a qual não possui evidências científicas comprovadas (VERLI, 2014).

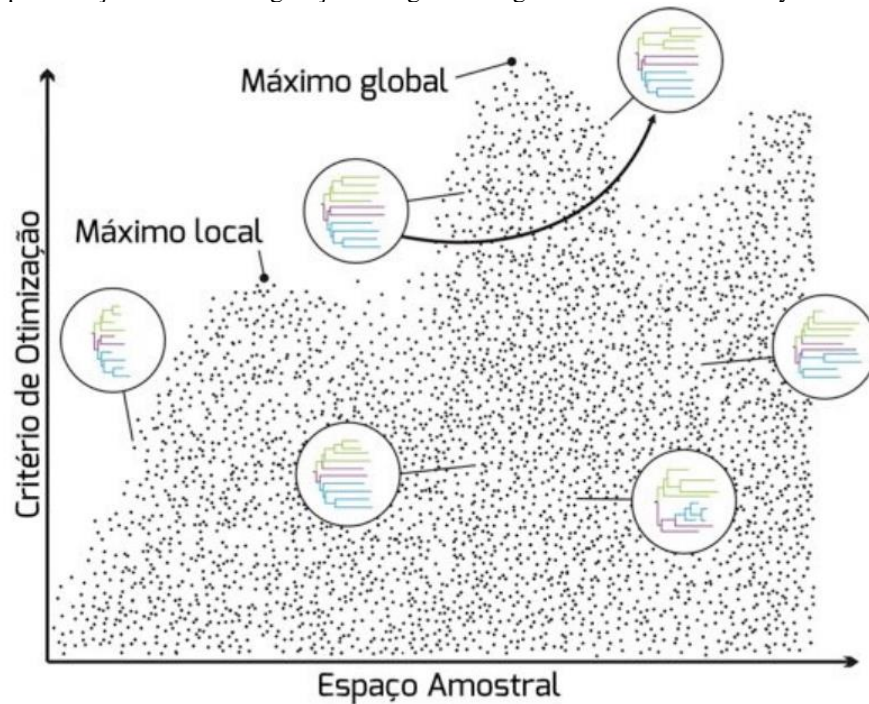
A inferência bayesiana aplica um modelo de máxima verossimilhança para geração das árvores, porém apresenta duas características que permitem o aumento da robustez das análises filogenéticas: a inserção de informações a priori e a geração de árvores com menor escore evolutivo. A inserção de informações a priori permite serem incluídas não só informações de modelos evolutivos (como ocorre na análise de máxima verossimilhança) mas também informações sobre grupos externos, relógio molecular, datações fósseis e outras informações acerca da natureza evolutiva das sequências analisadas (VERLI, 2014).

Durante o desenvolvimento do método Bayesiano, foi verificado que a aceitação apenas de escores maiores que o anterior, resultava em árvores amostradas em uma região de máximo local (Figura 5), e que quando escores menores que o anterior eram permitidos a árvore final correspondia à região amostrada em máximo global (Figura 5), dessa forma, caso um novo escore seja menor que o anterior, é gerado um número aleatório entre 0 e 1, se o resultado da divisão entre o escore novo e o escore antigo for maior que o número gerado aleatoriamente a nova geração “passa adiante” para uma nova modificação e subsequente novo cálculo de escore, caso o número seja menor que o valor gerado aleatoriamente, a geração é desconsiderada e novas modificações são realizadas na geração anterior. Este processo permitiu introduzir o fator de aleatoriedade dentro da análise evolutiva, culminando na geração de árvores com mais robustez estatística (ARCHIBALD; MORT; CRAWFORD, 2003).

2.6.2 Caracterização estrutural

Juntamente com análises filogenéticas, o entendimento do potencial funcional de uma região gênica é essencial para o entendimento do papel de determinada região dentro de um genoma. Diversos fatores podem ser observados para prever possíveis funções dos vírus endogenizados, tais como a conservação de domínios proteicos de suas proteínas preditas, estruturas tridimensionais das proteínas preditas e níveis de expressão de transcritos derivados dessas regiões, bem como a análise do contexto genômico em que estas regiões se encontram, como composição gênica da região flanqueadora, por exemplo.

Figura 5 - Representação amostral de gerações filogenéticas geradas em inferência bayesiana



Fonte: Verli, (2014).

Devido às limitações das estratégias baseadas em similaridade das sequências lineares das proteínas, em alguns casos a busca por domínios resulta em regiões sem anotação, devido à ausência de informações no banco de domínios ou à diferença entre as sequências a serem analisadas e as sequências que possuem anotação curadas nos bancos de dados. Entretanto, existem outros métodos que permitem a avaliação de domínios e estruturas proteicas conservadas com um maior grau de sensibilidade: a comparação de estruturas tridimensionais das proteínas. Visto que mesmo com a divergência de aminoácidos na estrutura linear, caso os aminoácidos divergentes possuam a mesma natureza físico-química podem resultar em dobramentos tridimensionais similares.

Somando todas as possíveis aplicações de metodologias computacionais para resolução de questões biológicas, a bioinformática é uma área essencial dentro da Paleovirologia, correspondendo às primeiras etapas de estudo envolvendo EVEs, onde ferramentas de busca por identidade permitem a identificação de sequências virais dentro dos genomas. A análise de domínios, motivos e estruturas moleculares de proteínas, codificadas por estes EVEs, permitem hipotetizar a função destes vírus dentro dos genomas.

Modelos evolutivos permitem identificar os momentos dos eventos de integração dos vírus no genoma de seus hospedeiros, além da relação evolutiva de EVEs e vírus atuais.

3 JUSTIFICATIVA:

A caracterização dos parasitas que coevoluem com o genoma hospedeiro é imprescindível para entender como ocorre o processo evolutivo entre estas entidades biológicas, uma vez que esse processo coevolutivo pode desencadear várias consequências tanto para a evolução do genoma hospedeiro como do parasita.

O estudo de vírus endógenos presentes nos genomas de culicídeos vetores trará dados valiosos para compreender quais vírus infectaram e/ou ainda infectam essas espécies. Além de gerar informações que podem ser cruzadas em estudos de metaviromica de culicídeos, permitindo entender se o material genético identificado nestes estudos advém de um vírus circulante ou de fragmentos de vírus endogenizado.

Dessa maneira, a caracterização dos EVEs vem a contribuir para o entendimento de questões evolutivas básicas e que poderão ser aplicadas no entendimento da relação vírus/vetor.

4 PERGUNTA CONDUTORA

Qual a real distribuição e os principais impactos dos Elementos Virais Endogenizados presentes nos genomas de Culicídeos?

5 HIPÓTESE

Os dados disponíveis da literatura refletem uma sub-representação da diversidade de Elementos Virais Endogenizados em culicídeos, e estes elementos podem exercer funções biológicas além do envolvimento em mecanismos de piRNAs.

6 OBJETIVOS

O presente trabalho possui um objetivo geral e três objetivos específicos.

6.1 Objetivo Geral

Entender a distribuição e o impacto de vírus endogenizados nos genomas de culicídeos disponíveis em bancos públicos e através de dados gerados no Instituto Aggeu Magalhães.

6.2 Objetivos Específicos

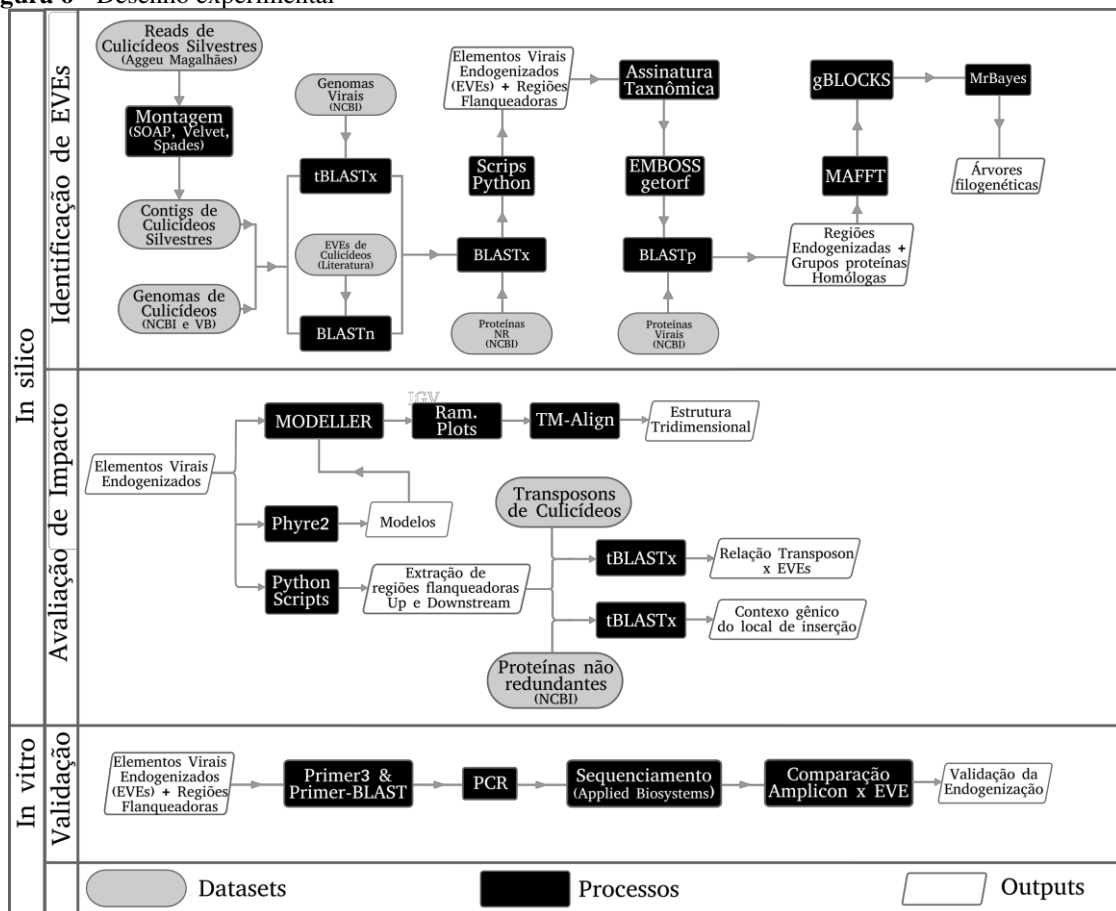
- a) Identificar os EVEs nos genomas de culicídeos quanto a família viral a que pertencem.
- b) Caracterizar os EVEs que potencialmente deram origem a novos produtos gênicos.
- c) Validar a endogenização de potenciais genes de origem viral *in vitro*.

7 METODOLOGIA

Atualmente, não existe uma convenção de nomenclatura para nomear elementos virais endogenizados, alguns autores utilizam a terminologia EVEs de forma a excluir os ERVs (do Inglês *Endogenous Retroviruses*). No presente trabalho, assim como em outros trabalhos da literatura (AIEWSAKUN; KATZOURAKIS, 2015; KATZOURAKIS; GIFFORD, 2010), usaremos EVEs para designar todos os grupos de vírus endógenos, incluindo os de origem retroviral.

O desenho experimental consiste em dois tipos de análises: *In silico* e *In vitro* (Figura 6). Primeiramente análises *in silico* permitiram a identificação de possíveis EVEs nos genomas de culicídeos bem como análises evolutivas das regiões endogenizadas. Após identificar EVEs que potencialmente possuem alguma função dentro dos genomas de culicídeos foram desenhados *primers*, seguido de análises *in vitro* para confirmar os resultados da primeira etapa por meio de PCR (do Inglês *Polymerase Chain Reaction*) e sequenciamento de Sanger.

Figura 6 - Desenho experimental



Fonte: O autor.

7.1 Obtenção de dados

Inicialmente foram definidas duas fontes de informação para o desenvolvimento do projeto, dados disponíveis em bancos de dados públicos, e dados gerados em projetos de colaboração no Instituto Aggeu Magalhães.

7.1.1 Datasets privados

Consiste em um *dataset* de 19 conjuntos de dados de sequenciamento genômico de baixa cobertura provenientes de 19 espécies de culicídeos (Tabela 2) gerados em um projeto de colaboração, desenvolvido durante os anos de 2017 e 2018 (SILVA, 2019). Os dados foram gerados através de sequenciamento em plataforma Illumina MiSeq, gerando conjuntos de *reads paired-end* de 75 nucleotídeos, de tamanhos distintos, provenientes de amostras coletadas em regiões de mata silvestre em Recife-PE e Sinop-MT. A qualidade dos dados foi verificada pela ferramenta FastQC v0.11.2 (ANDREWS, 2011) e os resultados foram concatenados através da ferramenta MultiQC (EWELS *et al.*, 2016).

Apenas o *dataset* de *Coquillettidia albicosta* (Cq_alb) apresentou *phred score* menor que 30 e regiões com conteúdo de adaptadores (Apêndice B), dessa forma, neste único *dataset* os dados foram tratados utilizando a ferramenta Trimmomatic v0.36 (BOLGER; LOHSE; USADEL, 2014) com os seguintes parâmetros: `trimmomatic-0.36.jar PE -threads 2 -phred33 arq_R1_001.fastq arq_R2_001.fastq -baseout arq ILLUMINACLIP:Trimmomatic-0.36/adapters/Nextera-PE-PE.fa:2:30:10 SLIDINGWINDOW:5:30`, o parâmetro `SLIDINGWINDOW:5: 30` foi utilizado para remover regiões que apresentaram uma qualidade menor que 30 (acurácia de base sequenciada menor que 99,9%), em janelas de 5 nucleotídeos.

Os *reads* foram montados utilizando as ferramentas *SoapDenovo2* v2.04 (LUO *et al.*, 2012), *VelvetOptimiser* v2.2.5 (ZERBINO; BIRNEY, 2008) e *Spades* v3.13.0 (BANKEVICH *et al.*, 2012) gerando então *contigs* (sequência consenso formada por sobreposição de *reads*) correspondentes às 19 espécies de culicídeos. Através da comparação de métricas como N50 (tamanho do menor contig em um conjunto de contigs que compõem 50% da montagem), número de contigs e tamanho de montagem proveniente destes três métodos (Apêndice C, Tabela S1), foram selecionadas as melhores montagens, que foram advindas da metodologia com a ferramenta SPADES (Apêndice C, Tabela S1), para análises

subsequentes. Devido à inexistência de informação acerca do tamanho do genoma destas espécies, não foi possível calcular a cobertura e a profundidade da montagem.

Tabela 2 - Genomas de Culicídeos gerados em projetos colaborativos sob coordenação do Instituto Aggeu Magalhães

Espécie	Código*	Montagem (pb)**	Nº. de Contigs	N50
<i>Aedomya squamipennis</i>	Ad_squa	1.775.830	5.997	585
<i>Aedes scapularis</i>	Ae_scap	1.896.643	4.765	510
<i>Aedes taeniorhynchus</i>	Ae_tae	7.085.622	15.574	532
<i>Coquillettidia arribalzagae</i>	Cq_arri	756.068	4.183	320
<i>Coquillettidia albicosta</i>	Cq_alb	386.029	856	834
<i>Coquillettidia chrysonotum</i>	Cq_chry	2.620.197	5.326	754
<i>Coquillettidia hermanoi</i>	Cq_herm	2.605.390	5.070	643
<i>Coquillettidia juxtamansonia</i>	Cq_juxt	2.341.348	5.466	641
<i>Coquillettidia venezuelensis</i>	Cq_vene	1.076.173	3.688	583
<i>Culex amazonensis</i>	Cu_amaz	2.955.160	8.024	331
<i>Culex corniger</i>	Cu_corn	462.878	2.074	432
<i>Culex nigripalpus</i>	Cu_nigr	623.991	3.393	315
<i>Limatus durhamii</i>	Li_durh	773.524	5.776	130
<i>Mansonia humeralis</i>	Ma_hume	2.508.905	4.838	716
<i>Mansonia titillans</i>	Ma_titi	4.849.885	8.830	662
<i>Mansonia wilsoni</i>	Ma_wils	2.179.545	17.307	116
<i>Psorophora cingulata</i>	Ps_cing	1.618.621	4.475	666
<i>Psorophora ferox</i>	Ps_ferox	435.193	1.964	335
<i>Trichoprosopom digitatum</i>	Tr_digi	1.442.868	7.797	343

Fonte: O autor.

Nota: * Código gerado para o presente estudo. ** Com base na comparação entre 3 métodos de montagem (Apêndice C).

7.1.2 Genomas disponíveis em bancos públicos

Durante o início deste projeto estavam disponíveis 37 genomas (27 espécie-específicos) de culicídeos nas plataformas NCBI e VectorBase. Após o *download* de cada genoma, cada sequência dos *scaffolds* (região consenso formada pela sobreposição de

contigs) foi marcada com uma “etiqueta” de acordo com a espécie e linhagem do culicídeo (Tabela 1).

Após a obtenção de cada *dataset*, algumas métricas foram analisadas, como o número de *scaffolds* ou de *contigs*, o N50 e o tamanho total da montagem.

7.2 Identificação de potenciais EVEs

A ferramenta BLAST (do Inglês *Basic Local Alignment Search Tool*) (ALTSCHUL, S. F. *et al.*, 1990) é reconhecida como o padrão ouro em análises de comparação de sequências, permitindo tanto a comparação de informação a nível de nucleotídeos (BLASTn e suas variantes), quanto a nível de aminoácidos (BLASTp e suas variantes).

Dessa forma, foram utilizados genomas virais depositados na plataforma GenBank do NCBI até a data de 12 de Janeiro de 2018, como sondas, contra os genomas de culicídeos, utilizados como banco de dados, em uma análise de tBLASTx, com os seguintes parâmetros: `tblastx -db mosquito_genome -query viral_query.fasta -out mosquito_genome.tblastx -outfmt 6 -num_threads 8 -max_target_seqs 1 -max_hsps 5 -evaluate 0,01`, onde tanto os genomas virais, quanto os genomas de culicídeos são traduzidos em ORFs para subseqüente comparação à nível de aminoácidos. Nesta etapa da análise, foi utilizado o parâmetro `-max_target_seqs 1` para recuperar o primeiro *hit* de cada sonda, ou seja, a primeira região do genoma do culicídeo que apresentou similaridade com proteínas virais, o parâmetro `max_hsps 5` para recuperar no máximo 5 *matches* por cada região específica de *query-subject*, o parâmetro `e-value 0,01` foi utilizado em todas as análises de alinhamento neste trabalho, visando reportar apenas *matches* com probabilidade de alinhamento por aleatoriedade menor que 0,01.

Para identificar outras cópias e/ou variantes de cada EVE, foi utilizada uma análise de tBLASTx, com os mesmos parâmetros anteriores, exceto o `max_target_seqs 1`, e com filtro de 80% de identidade utilizando cada EVE como sonda contra o genoma do culicídeo correspondente como banco.

Uma terceira análise de BLASTn foi utilizada, utilizando como sonda os EVEs de culicídeos disponíveis na literatura, contra os genomas dos culicídeos como banco, os parâmetros para esta etapa seguiram os mesmos das análises anteriores, sem a utilização do parâmetro `max_target_seqs`.

7.2.1 Removendo falsos positivos

Os *matches* (regiões dos genomas dos mosquitos que apresentaram similaridade com vírus) foram extraídos com *scripts* escritos em linguagem de programação Python e utilizados em uma análise de BLASTx utilizando contra o banco de proteínas não redundantes do NCBI (Com última atualização em 12 de Janeiro de 2018), com os mesmos parâmetros utilizados na análise anterior excluindo o parâmetro `-max_target_seqs`, para confirmar a origem viral da sequência, visto que proteínas do próprio hospedeiro, como DNAs e RNAs polimerases podem apresentar similaridade com polimerases virais.

Foram extraídos os 5 melhores *matches* de cada sonda, e foi aplicado a seguinte metodologia para definição das sequências como EVEs: caso a sonda retornar 4 *matches* ou menos, se o primeiro *match* for viral, a sonda é considerada um EVE, caso a sonda tenha pelo menos 5 *matches*, e destes 5 três forem virais, a sonda é considerada um EVE, este processo permite a eliminação de falsos positivos (onde o primeiro *match* é viral e os demais não) e falsos negativos (onde o primeiro *match* não é viral, e os demais são).

Devido à natureza evolutiva de vírus, transposons e vírus endógenos (KOONIN; DOLJA, 2014) as sequências que passaram nas etapas de filtro anteriores foram submetidas a uma análise de cruzamento com transposons de mosquitos, para isto foi utilizada uma análise de tBLASTx utilizando um banco tratado de transposons de culicídeos, ainda não publicado, criado em outro projeto de colaboração no Instituto Aggeu Magalhães.

7.2.2 Classificação Viral

Para classificação taxonômica foram utilizadas análises de BLASTx, com alta sensibilidade (parâmetro `word_size 2`), contra um banco de proteínas virais não redundantes, proveniente da plataforma NCBI, com última atualização em 05 de Fevereiro de 2019. As sequências que não apresentaram similaridade contra as proteínas deste banco foram utilizadas em uma análise de tBLASTx contra um banco de genomas virais não redundantes proveniente da plataforma NCBI com alta sensibilidade (parâmetro `word_size 2`), também com última atualização em 05 de fevereiro de 2019. Ambos os bancos foram confeccionados de modo que cada sequência do banco apresenta uma assinatura de nível taxonômico de acordo com os grupos taxonômicos presentes na plataforma ICTV com última atualização em 05 de fevereiro de 2019.

7.2.3 Obtenção de sequências proteicas preditas

Devido à divergência nucleotídica entre EVEs e vírus atuais, sequências proteicas foram recuperadas utilizando a ferramenta EMBOSS getorf (RICE; LONGDEN; BLEASBY, 2000), e utilizadas em uma análise de BLASTp contra as proteínas de cada vírus atual correspondente, visto que diferentes regiões dos vírus podem ser endogenizadas.

Após esta etapa, as análises subsequentes foram realizadas apenas para os EVES da família *Chuviridae*, devido ao padrão de regiões endogenizadas identificados.

7.3 Predição e função de contexto genômico

Para entender os processos evolutivos relacionados a eventos de endogenização viral, é necessário entender o contexto genômico que estes elementos se encontram.

7.3.1 Estrutura das regiões flanqueadoras

Para avaliar a composição das regiões flanqueadoras, foram utilizados scripts em linguagem Python para extração e recuperação das sequências de 10.000 nucleotídeos *upstream* e *downstream* de cada EVE analisado. Foram avaliados dois fatores: a presença de transposons e de genes característicos do genoma de mosquitos. Esta etapa foi realizada apenas para os genomas recuperados de bancos públicos.

Vale ressaltar que apenas sequências que possuem regiões “avaliáveis” (com pelo menos 10.000 nucleotídeos em cada flanco e sem conteúdo de Ns) foram selecionadas para análises evolutivas e de possível função.

7.3.1.1 Transposons

Devido ao fato de transposons ativos serem um importante fator para integração viral, foi avaliada a composição de elementos transponíveis presentes nas regiões flanqueadoras aos vírus endógenos, a análise foi realizada através de busca por similaridade utilizando a ferramenta BLASTn contra mesmo banco utilizado na etapa de filtragem dos vírus endógenos bem como o banco de transposons da plataforma RepBase (JURKA, 1999).

7.3.1.2 Genes

Para identificar a presença de genes, as regiões flanqueadoras que não apresentarem transposons ou resquícios virais em sua estrutura, foram utilizadas em uma busca por similaridade via BLASTn no banco VectorBase, visando identificar a presença de genes dos próprios culicídeos nestas regiões.

7.3.2 Alinhamento e obtenção de blocos conservados

Alinhamentos múltiplos locais através da ferramenta MAFFT (KATOHI *et al.*, 2001) foram realizados para dois grupos de proteínas, utilizando as glicoproteínas preditas dos EVEs da família *Chuviridae* e suas possíveis correspondentes atuais identificadas na análise anterior e para poliproteínas de transposons da superfamília BEL-Pao identificados nas regiões flanqueadoras. Cada alinhamento foi realizado de acordo com as estruturas a serem alinhadas, em caso de estruturas com apenas um domínio conservado foi utilizado o modelo L-INS-i, enquanto estruturas com mais de um domínio conservado (poliproteínas, por exemplo) foi utilizado o modelo E-INS-i. A matriz de distância para cada alinhamento foi a BLOSUM45 devido a divergência de proteínas de EVEs e proteínas de vírus atuais. A exclusão de regiões variáveis e subsequente obtenção de blocos conservados foi realizada utilizando a ferramenta Gblocks (CASTRESANA, 2000) com parâmetros relaxados (permissão de *gaps* e validação de blocos com *indels*) visto a possibilidade de divergência entre as sequências mesmo em nível protéico.

7.3.3 Reconstruções evolutivas

O melhor modelo de substituição de aminoácidos para cada filogenia foi avaliado pela ferramenta SMS (LEFORT; LONGUEVILLE; GASCUEL, 2017) presente na plataforma online ATCG (<http://www.atgc-montpellier.fr/>).

A ferramenta MrBayes v3.2.2 (RONQUIST; HUELSENBECK; TESLENKO, 2010) foi utilizada para reconstrução de árvores bayesianas, utilizando gerações suficientes para que o desvio padrão entre os suportes de ramo das diferentes gerações fosse igual ou menor que 0,05. Ao final de cada análise foi realizado um *burn in* das 25% primeiras gerações, e foi gerada uma árvore consenso com as gerações restantes.

7.3.4 Modelagem comparativa

Modelagens moleculares foram realizadas para glicoproteínas de EVEs da família *Chuviridae* identificadas em três diferentes grupos: I - Solo-glicoproteínas, que correspondem a glicoproteínas sem assinatura de transposons em seus flancos; II glicoproteínas relacionadas à transposons e III - glicoproteínas de Chuvirus verdadeiros, anteriormente identificados em mosquitos. Para as glicoproteínas do grupo II, uma matriz de distância de alinhamento de aminoácidos foi criada com o *software* UGENE (OKONECHNIKOV; GOLOSOVA; FURSOV, 2012), uma sequência aleatório de cada agrupamento com mais de 90% de similaridade foi escolhida (Apêndice C, Tabelas S2 e S3).

O servidor Phyre2 (KELLEY *et al.*, 2015) foi utilizado para selecionar o modelo para cada proteínas de interesse. Cada modelo selecionado teve sua estrutura 3D recuperada do *Protein Data Bank* (PDB) (Berman *et al.* 2000), e foram submetidos junto com a sequência de aminoácidos de cada glicoproteína de interesse ao pacote MODELLER versão 9.23. Os modelos preditos foram avaliados pelos seus parâmetros estereoquímicos utilizando a ferramenta Procheck (Laskowski *et al.* 1993), além disso, foram realizados alinhamentos de estruturas tridimensionais, utilizando a ferramenta TM-Align (ZHANG; SKOLNICK, 2005) entre os modelos de cada um dos três grupos de glicoproteínas.

7.4 Análises *in vitro*

EVEs que, após predições computacionais, foram identificados como possíveis novos produtos gênicos no genoma de mosquitos das espécies *Aedes aegypti*, *Aedes albopictus* ou *Culex quinquefasciatus*, foram selecionados para etapa de validação em bancada. Estas três espécies foram escolhidas devido à disponibilidade de espécimes advindos do insetário presente no Instituto Aggeu Magalhães, bem como advindas de coletas de campo. Para validação *primers* foram desenhados utilizando a ferramenta Primer3 (HANCOCK, JOHN M., 2004) e validados utilizando a ferramenta PrimerBLAST (YE *et al.*, 2011) da plataforma NCBI. Os *primers* foram desenhados com o intuito de amplificar um fragmento que contenha uma porção do genoma viral, e outra porção do genoma do culicídeo (Figura 7).

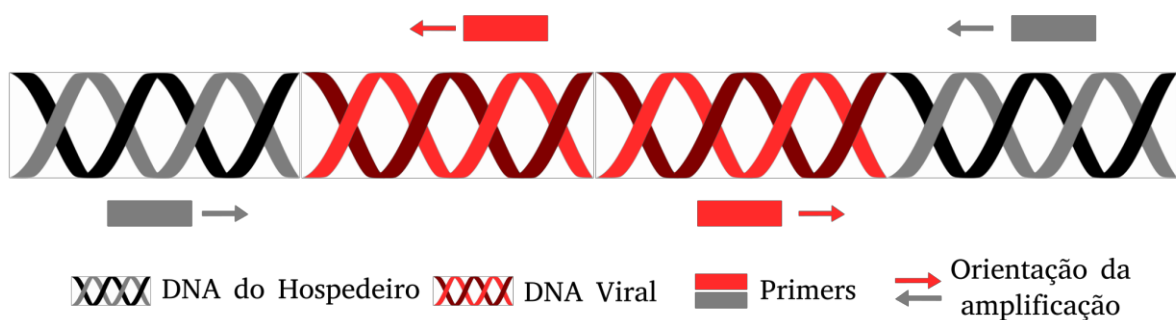
Informações de cada par de *primers* estão disponíveis no (Apêndice C, Tabela S4). As Reações em Cadeia da Polimerase (PCRs, do Inglês Polymerase Chain Reactions) foram realizadas com o kit GoTaq-Flexi G2 DNA-polymerase usando o protocolo do fabricante (Promega ©). Todas as PCRs foram conduzidas com uma solução final de 25 uL contendo 1

uL de cada *primer* (10 uM), 2 uL de dNTPs (0,2 mM de cada nucleotídeo), 5 uL de GoTaq Flexi Buffer, 4 uL de MgCL2 (25 mM), 0,25 uL de GoTaq-Flexi G2 DNA-Polymerase, 3 uL de amostra de DNA (normalizado para concentração de 10 ng/uL) e 8,75 uL de água ultrapura.

As ampliações foram conduzidas com o mesmo protocolo: Desnaturação inicial em 94 °C por 2 minutos, 45 ciclos de: desnaturação em 94°C por 1 minuto, anelamento em 51,8~58,3 °C (dependendo da temperatura de anelamento de cada primer) por 50 segundos e extensão em 72 °C durante 1 minuto com uma extensão ao final dos 45 ciclos em 72 °C por 5 minutos. Os produtos de PCR foram visualizados em gel de agarose 2%, as reações com bandas inespecíficas foram purificadas utilizando kit de beads magnéticas (Agencourt AMPure XP, Beckman Coulter Life Sciences ©), e as bandas de tamanho esperado foram sequenciadas por sequenciamento Sanger através da plataforma DNA ABI Prism 3100 Genetic Analyser (Applied Biosystems) para ambas as fitas (direta e reversa).

Os eletroferogramas foram analisados com o *software* Geneious Prime versão 2019.1.3 (KEARSE *et al.*, 2012) e as regiões consenso foram geradas por alinhamento global par-a-par e comparadas com as regiões esperadas através da ferramenta Aliview (LARSSON, 2014).

Figura 7 - Estratégia de desenho de primers para os vírus endógenos



Fonte: O autor.

7.5 Visualização de dados

As árvores filogenéticas foram representadas e anotadas utilizando a ferramenta iTOL (LETUNIC; BORK, 2007), os gráficos foram construídos utilizando as bibliotecas Seaborn (<https://seaborn.pydata.org/>) e Matplotlib (<https://matplotlib.org/>) da linguagem de programação Python, os mapas genéticos foram construídos utilizando o pacote em R

karyoploteR (GEL; SERRA, 2017) e todos os detalhes finais das figuras foram ajustados usando a ferramenta Inkscape 0.92.3 (<https://inkscape.org/>).

8 RESULTADOS

Eventos de mutações ao longo da história evolutiva do genoma hospedeiro podem gerar mudanças de fase de leitura na sequência inicialmente endogenizada, Palatini e seus colaboradores (PALATINI *et al.*, 2017) utilizaram em seu estudo um limiar de 100 nucleotídeos entre um EVE e outro, porém Whitfield e seus colaboradores (WHITFIELD *et al.*, 2017) demonstram que EVEs comumente estão inseridos em agrupamentos, inclusive sobrepostos, dessa forma, para uma contagem inicial, não foi estabelecido nenhum limiar de distância entre os EVEs, totalizando 1998 sequências identificadas como potenciais EVEs.

8.1 Diversidade de EVEs em culicídeos: Origem de hospedeiros, distribuição em culicídeos e regiões endogenizadas

Das 1998 sequências como potenciais vírus endógenos, 1842 foram classificadas a nível de família ou gênero (no caso de inexistência do nível de família) viral e 156 foram agrupadas no grupo de vírus não classificados, devido à inexistência de assinatura taxonômica de vírus conhecidos (Figura 8 A).

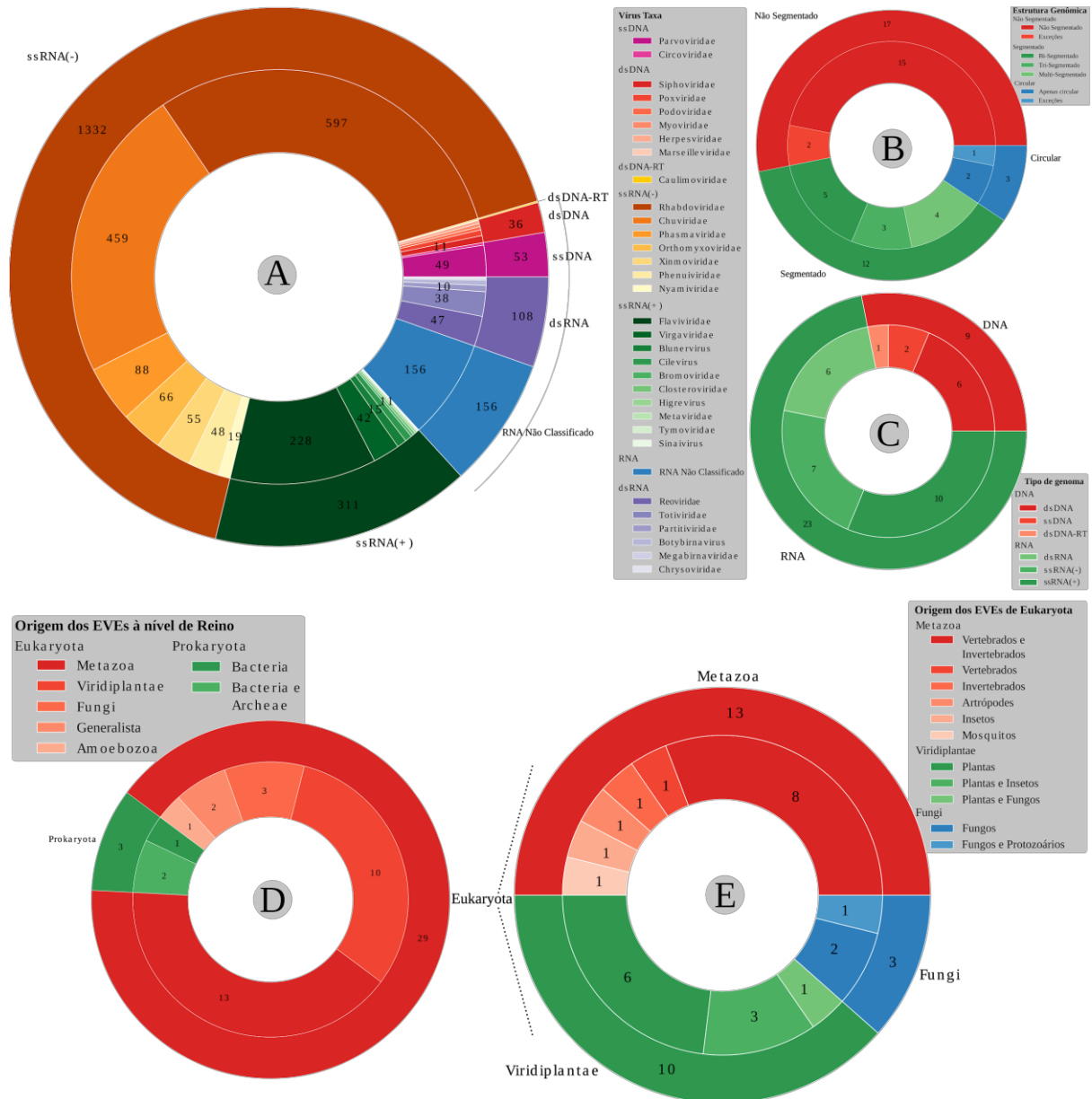
Analisando a diversidade de famílias e gêneros, foi possível identificar 32 taxa virais, onde os EVEs derivados de vírus de ssRNA (sentido positivo ou negativo) compõem a maior parcela de elementos identificados, totalizando mais de 1600 sequências. Além de EVEs derivados de vírus de RNA, também foi possível identificar cerca de 90 elementos derivados de vírus de DNA (Figura 8 A).

Avaliando diversos fatores relacionados ao número de famílias virais identificadas, foi possível identificar famílias virais das três categorias de estrutura genômica viral (segmentado, não segmentado e circular), onde a maioria das famílias virais de EVEs identificados são de vírus não segmentados (Figura 8 B). Também foi possível avaliar que tanto a quantidade de EVEs quanto o número de famílias e gêneros pertencentes à vírus de genoma de RNA (Figura 8 C).

Verificando os principais hospedeiros de cada família viral que possui representantes endogenizados em culicídeos, foi possível identificar tanto vírus de origem eucariótica quanto procariótica (Figura 8 D) bem como uma diversidade de taxa eucarióticos que não se

limita apenas a famílias virais de animais (Metazoa), mas também de plantas e fungos, totalizando 13 famílias virais que possuem como seu hospedeiro principal organismos pertencentes aos reinos Viridiplantae e Fungi (Figura 8 E).

Figura 8 - Diversidade de famílias virais, estrutura genômica e hospedeiros de origem de EVEs identificados em genomas de culicídeos



Fonte: O autor.

Legenda: **A:** Número total de EVEs distribuídos por: Classes virais segundo classificação de Baltimore (círculo externo); Taxa virais seguindo o sistema taxonômico proposto pelo ICTV (círculo interno); **B:** Número de famílias virais por tipo de estrutura genômica; **C:** Número de famílias virais por natureza de material genético que compõe o genoma viral; **D:** Número de famílias virais distribuídas por hospedeiros principais a nível de presença ou ausência de carioteca; **E:** Número de famílias virais distribuídas por hospedeiros principais dentro do super reino Eukaryota. Informações utilizadas para construção dos gráficos se encontram no Apêndice C Tabela S5.

Levando em consideração a diversidade de genomas analisados, foi possível identificar que os culicídeos do gênero *Aedes* possuem uma maior quantidade de EVEs em seus genomas, sendo a montagem com maior número de elementos a montagem *Aedes aegypti* Aag2, sequenciada por abordagem de *long-read* (WHITFIELD *et al.*, 2017), e que essa quantidade varia de acordo com a montagem do genoma analisada (Figura 9 A). Também foi possível verificar apenas uma pequena quantidade de EVEs proveniente dos *datasets* privados (Inseridos majoritariamente na categoria Outros, Figura 9 A), provavelmente devido a incapacidade de resolução de regiões repetitivas com genoma de baixa cobertura, o que pode enviesar a anotação de EVEs (WHITFIELD *et al.*, 2017).

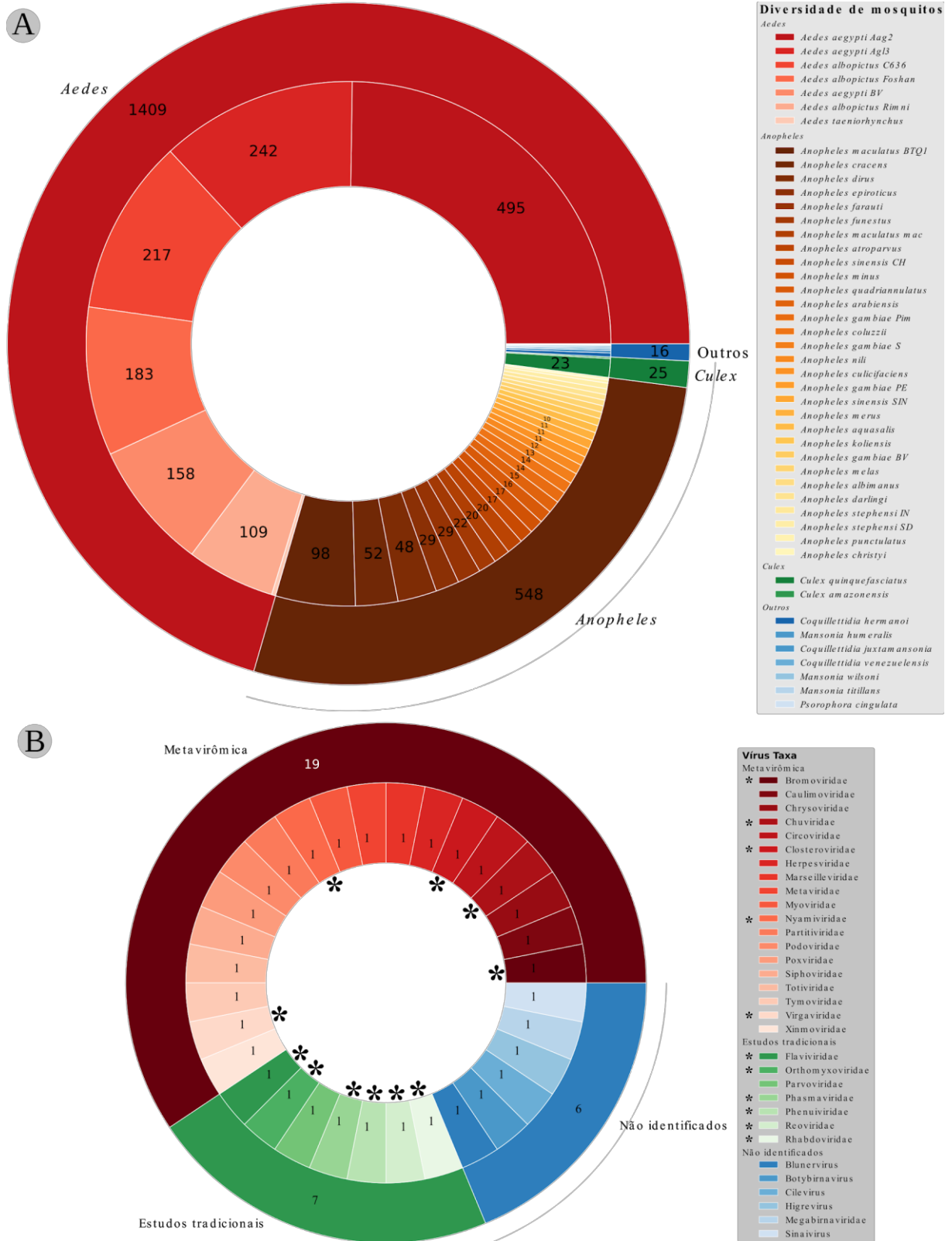
Cruzando as informações levantadas nestas análises iniciais, com informações da literatura sobre: Vírus identificados em culicídeos com abordagens tradicionais (Identificação, amplificação e sequenciamento do genoma viral), com abordagens de metavirômica (Identificação de material genético viral por sequenciamento de RNA total de amostras biológicas) e famílias virais previamente identificadas como EVEs em culicídeos foi possível identificar alguns pontos interessantes (Figura 9B):

- a) Dos 32 taxa virais de EVEs identificados, 19 possuem representantes identificados em culicídeos por estudos de metavirômica; 7 identificados em estudos tradicionais e 6 ainda não tendo representantes identificados em culicídeos;
- b) Dos 32 taxa virais de EVEs identificadas neste estudo, apenas 11 já haviam sido descritas em outros estudos de EVEs em culicídeos, representando uma diversidade de 21 novas famílias de EVEs em culicídeos.

Avaliando o panorama de distribuição de família/gêneros de EVEs com os genomas analisados (Figura 10), foi possível identificar uma distribuição desigual não só em número de elementos, mas também em diversidade de famílias quando comparadas duas montagens de uma mesma espécie, como *Aedes aegypti* Aag2 e *Aedes aegypti* BV, *Anopheles maculatus* BTQ1 e *Anopheles maculatus mac* e nas 4 montagens de *Anopheles gambiae*, por exemplo.

Também foi possível identificar a presença de diversos EVEs pertencentes à categoria de RNA não classificados e EVEs das famílias *Rhabdoviridae* e *Chuviridae* em quase todos os genomas analisados (Figura 10).

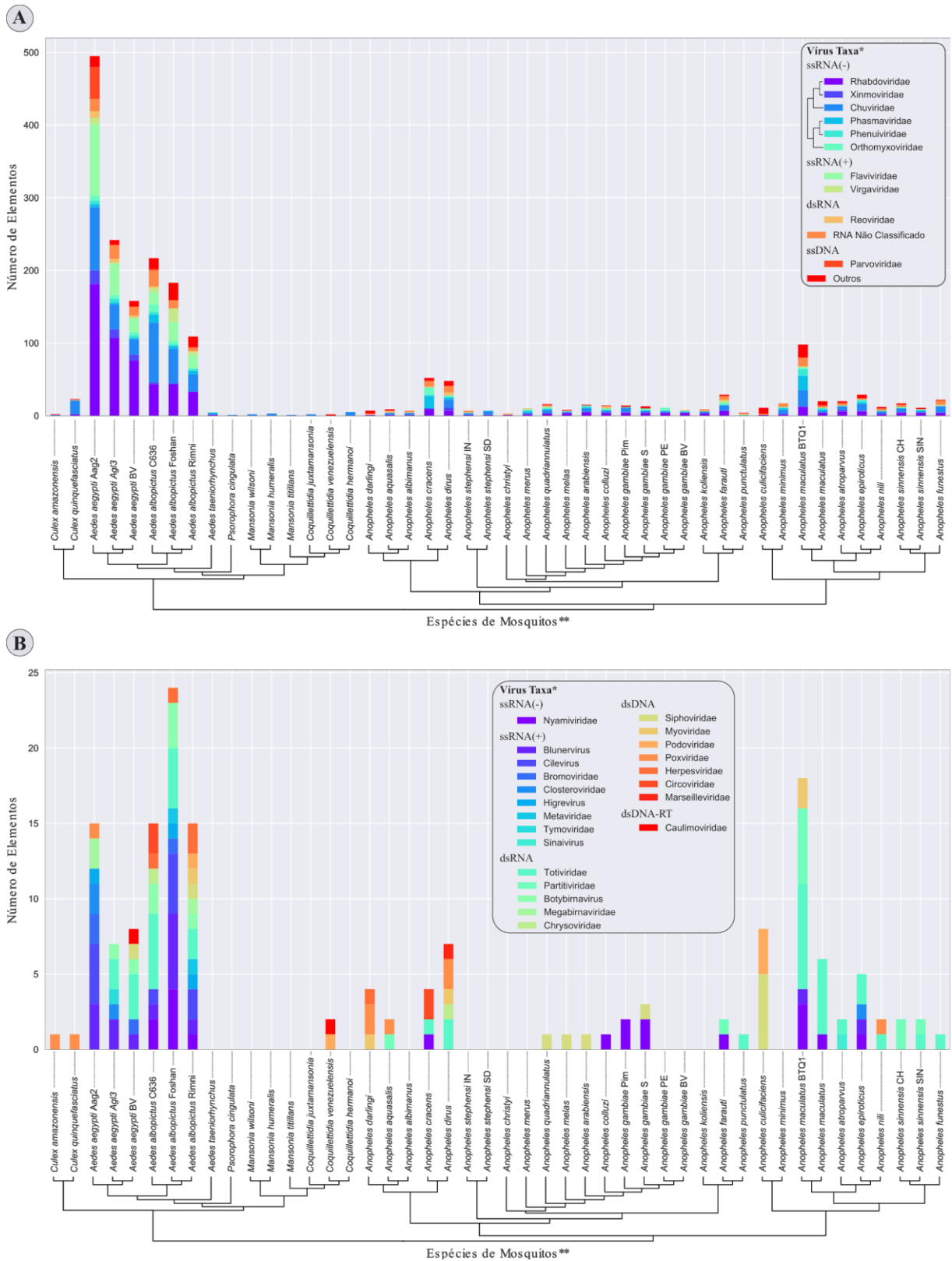
Figura 9 - Diversidade de elementos por espécies de culicídeo e comparação com dados da literatura.



Fonte: O autor.

Legenda: **A:** Número de elementos por espécie/linhagem de culicídeo. **B:** Comparativo com dados da literatura sobre identificação de famílias virais em culicídeos, * representam famílias virais já identificadas como EVEs de culicídeos por outros estudos. Informações utilizadas para construção dos gráficos se encontram no Apêndice C.

Figura 10 - Distribuição de EVEs por espécie/linhagem de culicídeo



Fonte: O autor.

Legenda: **A**: Distribuição de taxa que apresentaram mais de 40 elementos por táxon; **B**: Distribuição de taxa que apresentam menos de 40 elementos por táxon. * Taxonomia seguindo o sistema de classificação do ICTV; ** Distribuição de espécies definidas por uma Análise filogenética de inferência bayesiana utilizando regiões nucleotídicas do gene Citocromo Oxidase I, com 1 milhão de gerações a partir de 3 árvores iniciais.

Avaliando a estrutura das regiões endogenizadas, foi possível definir a estrutura proteica de 1889 das 1998 sequências como potenciais vírus endógenos, onde foi possível identificar a maioria das sequências incluídas em grupos de proteínas provenientes de genes “*hallmark*” (Figura 11) como glicoproteínas, nucleoproteínas e polimerases (KOONIN; DOLJA, 2014).

Também foi possível verificar a diferença no número das regiões endogenizadas por família viral, como por exemplo 239 nucleoproteínas, 133 polimerases e 132 glicoproteínas endogenizadas da família *Rhabdoviridae* e 440 glicoproteínas, 18 polimerases e uma nucleoproteína endogenizadas da família *Chuviridae*.

Devido a presença quase ubíqua nos genomas analisados de EVEs da família *Chuviridae*, bem como um padrão desigual de endogenização (Figura 11) e composição de transposons em seus flancos, quando comparado com EVEs das famílias *Rhabdoviridae* e *Flaviviridae* (WHITFIELD *et al.*, 2017), essa família foi selecionada para prosseguimento das análises.

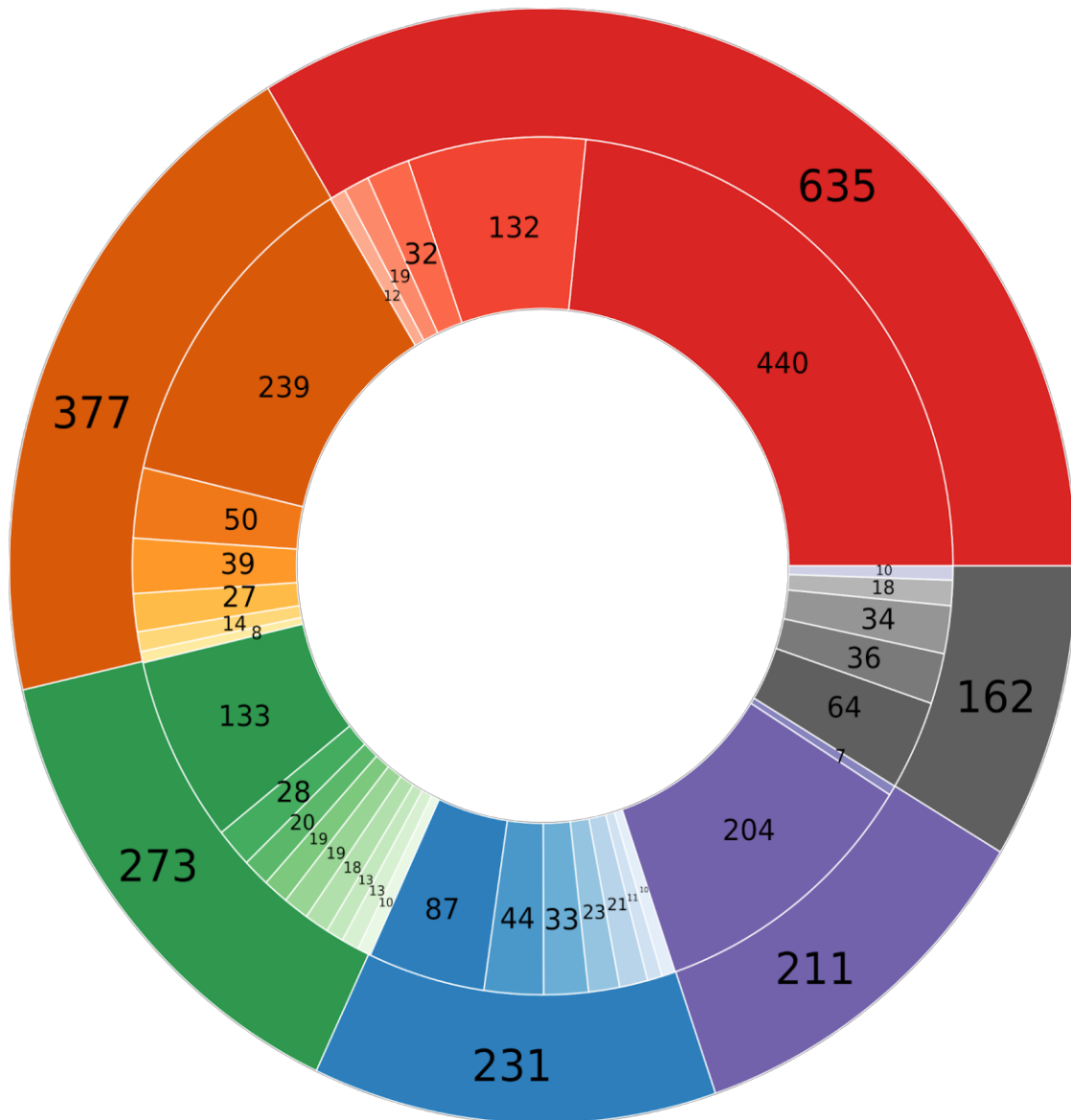
8.2 Família *Chuviridae*

Avaliando o resultado da etapa de assinatura taxonômica e das regiões endogenizadas, os 459 elementos assinados como membros da família *Chuviridae* apresentaram 13 diferentes *hits*, destes 9 correspondem à glicoproteínas, 3 à RNAs-polimerases dependentes de RNA e 1 corresponde à nucleoproteína (Apêndice C, Tabela S6).

Quando contabilizadas o número de sequências para cada estrutura, temos 440 sequências que apresentaram similaridade com glicoproteínas, 18 com RNAs-polimerases dependentes de RNA e 1 com nucleoproteína. Selecionamos então as glicoproteínas para análise, foi realizado um processo de curagem manual em 3 etapas (informações disponíveis no Apêndice C, Tabela S6):

- a) Remoção de possíveis sequências redundantes: EVEs de um mesmo genoma, que apresentaram 100 nucleotídeos de distância entre si, com similaridade para diferentes regiões de uma mesmo glicoproteína de *Chuvirus* foram considerados como a mesma sequência, onde a sequência mais próxima da extremidade 3' foi considerada. Esta etapa removeu 16 sequências, viabilizando 424 sequências para análises posteriores;

Figura 11 - Distribuição de EVEs por grupo de proteínas endogenizadas



Taxon por termo

Glycoprotein

- Chuviridae
- Rhabdoviridae
- Phasmavaridae
- Xinmoviridae
- Outros

Nucleoprotein

- Rhabdoviridae
- Phasmaviridae
- Phenuiviridae
- Orthomyxoviridae
- Totiviridae
- Outros

Polymerase

- Rhabdoviridae
- Orthomyxoviridae
- Totiviridae
- Others
- Nyamiviridae
- Chuviridae
- Blunervirus
- Unclash
- Partitiviridae

Others

- Outros
- Parvoviridae

- Virgaviridae
- Orthomyxoviridae
- Xinmoviridae
- Rhabdoviridae
- Cilevirus
- Polyprotein*
- Flaviviridae
- Outros
- Hypothetical*
- Unclash
- Rhabdoviridae
- Reoviridae
- Outros
- Siphoviridae

Fonte: O autor.

- a) Remoção das sequências que apresentaram regiões protéicas conservadas com menos de 100 aminoácidos, devido a impossibilidade de realizar análises de alinhamento ou filogenéticas com outras glicoproteínas de Chuvirus, que apresentam tamanho médio de 600 aminoácidos. Esta etapa removeu 37 sequências, restando 387 sequências para análises posteriores;
- b) Remoção de sequências que não apresentaram pelo menos 4 mil nucleotídeos de região flanqueadora, seja pelo tamanho do *contig/scaffold*, pela proximidade ao final ou início do *contig/scaffold* ou pela proximidade à regiões de N. Esta etapa removeu 108 sequências, viabilizando 279 sequências para análises posteriores.

Após a aplicação dos filtros, as regiões protéicas das 279 sequências apresentaram 2 padrões de estruturas, um dos padrões apresentou 241 sequências com blocos conservados e foi utilizado nas análises posteriores.

8.2.1 Análises de regiões flanqueadoras:

Foram verificadas as estruturas das regiões flanqueadoras, onde foi identificada a presença de poliproteínas com assinatura dos domínios de Protease, Transcriptase Reversa, RNaseH e Integrase, que representam estruturas de Elementos Transponíveis da superfamília BEL-Pao. Os transposons ativos da superfamília BEL-Pao possuem uma poliproteína com esses quatro domínios, uma segunda ORF que codifica uma proteína Gag além de regiões LTR em seus flancos, que servem como sinal de reconhecimento das proteínas envolvidas na transposição, regiões PBS (*Primer Binding Site*) e PPT (*Polypurine Tract*) essenciais para a retrotranscrição e início da síntese do DNA do retrotransposon, respectivamente. Eventualmente proteínas similares à proteína de envelope viral podem compor a última região proteica do elemento (LLORENS *et al.*, 2011).

Devido a este achado, buscamos avaliar se as glicoproteínas de Chuvirus encontradas endogenizadas pertenciam ou não a elementos BEL-Pao (fazem parte da estrutura dos elementos BEL como uma glicoproteína que possivelmente dá origem ao envelope) e se o alto número de cópias encontradas em alguns genomas era devido a replicação da glicoproteína junto com esses elementos transponíveis. Para tal, avaliamos a distribuição destes elementos transponíveis com ou sem glicoproteínas em sua estrutura e caracterizamos as regiões LTR com a ferramenta LTR_FINDER (XU; WANG, 2007).

Foi possível identificar 3 contextos genômicos das glicoproteínas endogenizadas (Figura 12): Apenas a glicoproteína, sem assinatura de Retroelementos BEL-Pao em suas regiões flangeadoras (Figura 12 A); A glicoproteína flangeada por estruturas incompletas de Retroelementos BEL-Pao (Figura 12 B) e por fim, a glicoproteína fazendo parte de Retroelementos BEL-Pao completos (Figura 12 C).

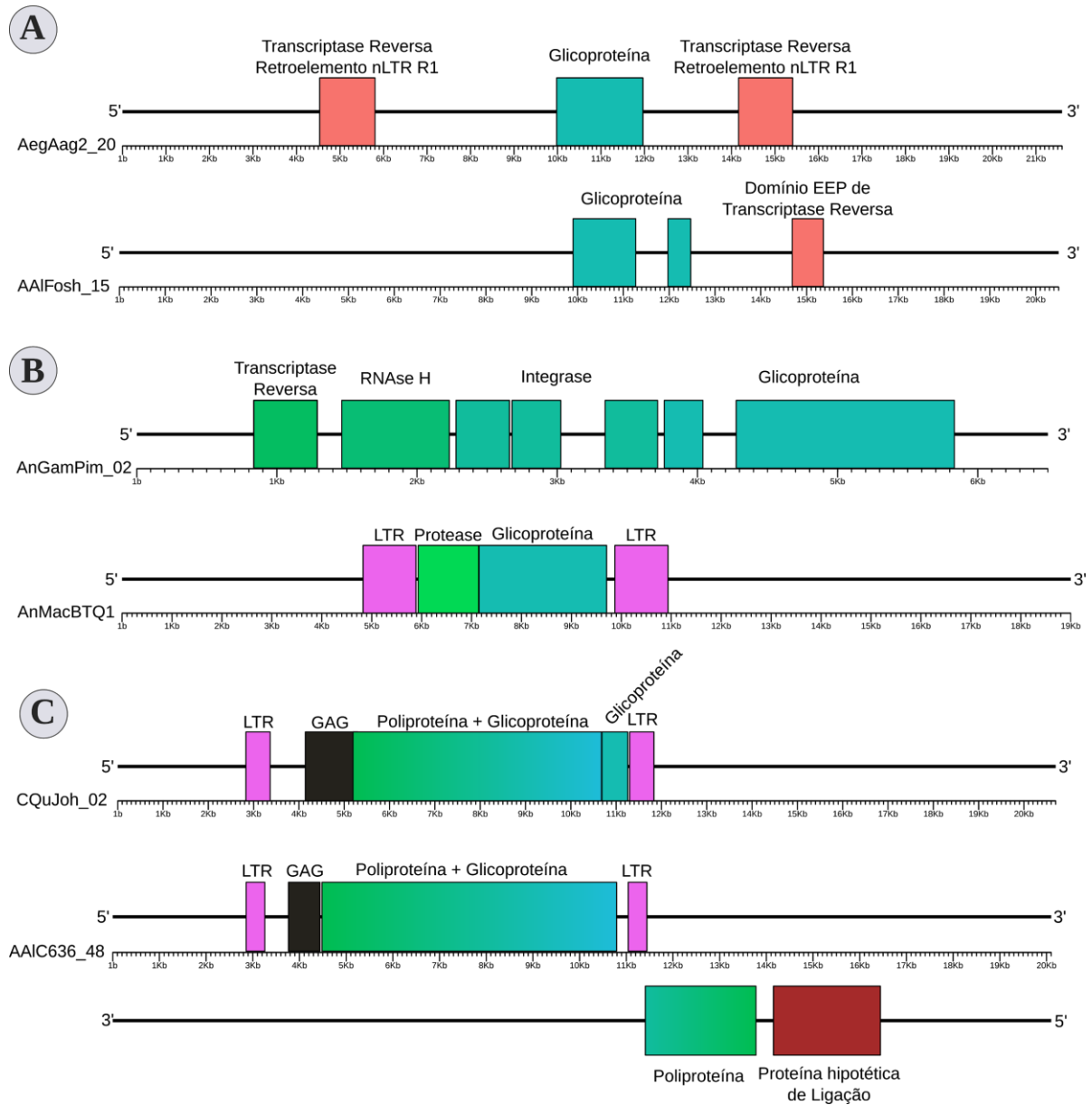
Das 241 glicoproteínas analisadas, apenas 36 não apresentaram assinatura de Retroelementos BEL-Pao em seus flancos, 11 parecem fazer parte de Retroelementos BEL-Pao completos e 194 em Retroelementos BEL-Pao incompletos. Informações sobre anotação tanto das glicoproteínas quanto de suas regiões flangeadoras estão disponíveis no Apêndice C, Tabela S5.

8.2.2 Modelagem molecular

Para etapas de modelagem molecular foram selecionadas as glicoproteínas dos *Chuvirus* de mosquitos, tanto a linhagem chinesa quanto norte-americana, bem como glicoproteínas endogenizadas, apresentando ou não estruturas de Retroelementos BEL-Pao em seus flancos, os modelos selecionados pela ferramenta Phyre bem como a porcentagem de resíduos presentes em cada uma das regiões do gráfico de Ramachandran estão disponíveis na Tabela 3.

Após etapas de modelagem, foi possível identificar que tanto as glicoproteínas de vírus circulantes, quanto as glicoproteínas endogenizadas, relacionadas ou não à retroelementos BEL-Pao, possuem uma estrutura molecular tridimensional com o mesmo dobramento, característica indicada pelos valores de *TM scores* superiores a 0,7 (Figura 13) uma vez que *scores* maiores que 0,5 são considerados por resultarem no mesmo dobramento tridimensional das proteínas (ZHANG; SKOLNICK, 2005).

Figura 12 - Exemplos de glicoproteínas de Chuvirus endogenizados em diferentes contextos genômicos



Fonte: O autor.

Legenda: **A**: Sem a presença de Retroelementos BEL; **B**: Com a presença de Retroelementos BEL-Pao não-autônomos (ausência de maquinaria completa); **C**: Fazendo parte de Retroelementos BEL-Pao autônomos (Com a maquinaria completa para transposição).

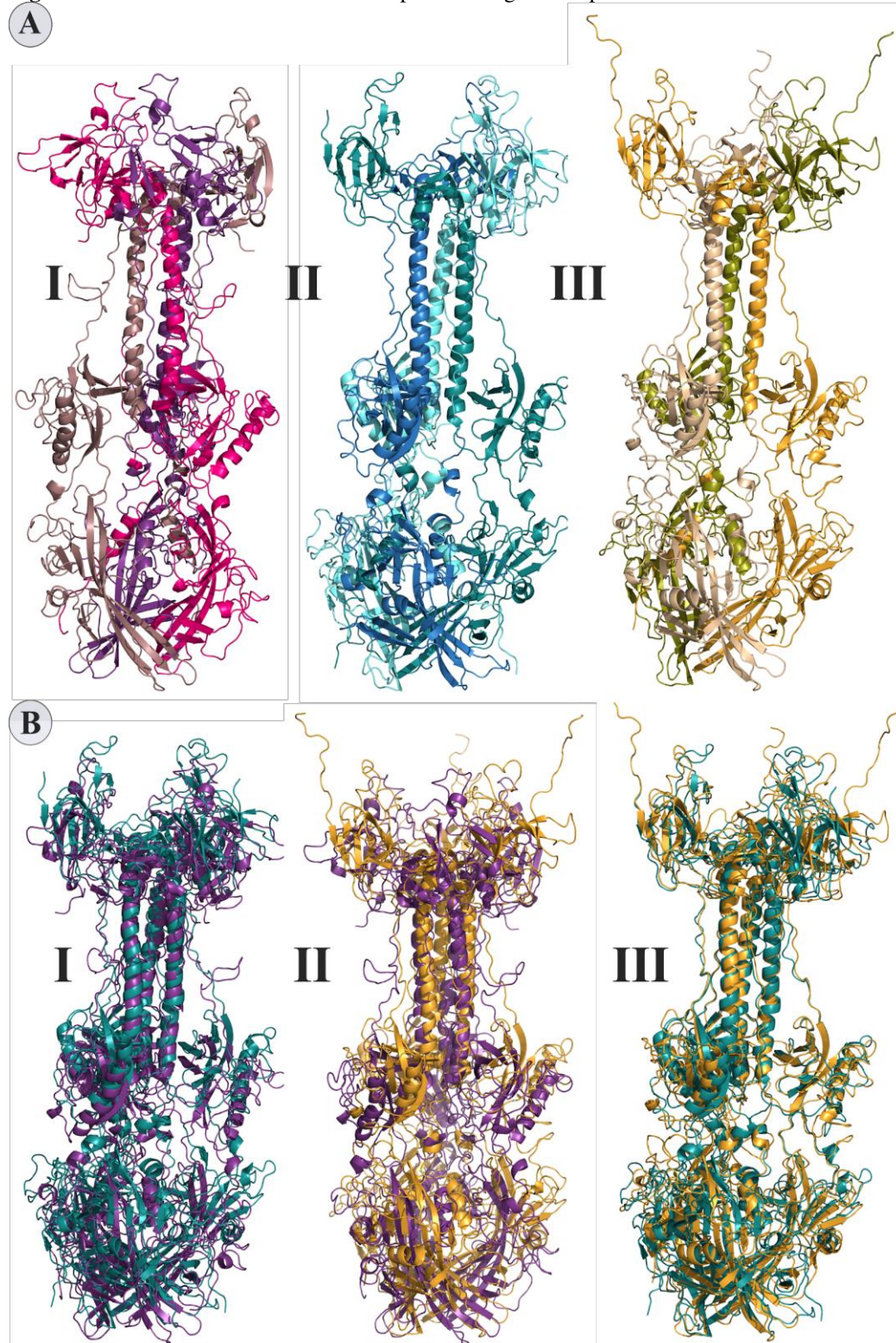
Tabela 3 - Glicoproteínas submetidas à processos de modelagem molecular

Glicoproteína	Modelo*	Template	Core**	Allow**	Gener**	Disall**
AAeBV_JXPU01223002.1_66_2032	3	6ESC	84,2	12,8	2,3	0,7
AAIC636_12	3	6ESC	83,7	13,5	2,2	0,6
AAIC636_23	1	6ESC	85,6	12,0	2,1	0,3
AAIFosh_15	3	6ESC	85,9	11,5	2,2	0,5
AegAag2_45	4	6ESC	82,3	14,2	2,6	0,9
AegAag2_20	3	6ESC	85,1	12,3	1,7	0,8
AnEpiE_KB670167_662_1045	5	6ESC	84,6	12,2	2,9	0,3
CQuJoh_07	4	6ESC	85,3	11,7	2,1	0,9
CQuJoh_09	2	6ESC	83,2	13,6	2,5	0,7
Mos8Chu0	3	6ESC	83,2	13,8	2,2	0,8
AAIC636_04	1	6ESC	87,1	10,9	1,1	0,9
AAIC636_15	1	6ESC	86,4	10,9	1,9	0,8
AnGamPim_05	4	6ESC	85,4	11,0	2,9	0,7
AnMacBtQ1_01	5	6ESC	83,1	13,0	2,8	1,1
BEL-199	5	5V2S	86,7	11,4	1,4	0,5
CQuJoh_03	4	6BM8	87,0	11,0	1,4	0,6
AAIFosh_21	2	5V2S	88,3	9,3	1,6	0,8
Wuhan Mosquito Virus 8	5	3NW8	83,2	12,6	2,9	1,3

Fonte: O autor.

Nota: * Melhor modelo de acordo com a ferramenta Procheck, ** Informações relacionadas ao gráfico de Ramachandran.

Figura 13 - Modelos moleculares obtidos por modelagem comparativa



Fonte: O autor.

Legenda: **A:** Modelos tridimensionais onde: **I** elemento AegAag2_20, um exemplo de solo-glicoproteína; **II** elemento AA1C636_23, um exemplo de glicoproteína fusionada à um retroelemento Bel-Pao completo; e **III** representa a glicoproteína de Mos8Chu0, um Chuvirus circulante. **B:** Representa os alinhamentos das estruturas tridimensionais entre: **I:** AegAag2_20 em roxo e AA1C636_23 em azul, com TM-score igual a 0,79415; **II:** AegAag2_20 em roxo e Mos8Chu0 em amarelo, com TM-score igual a 0,76639; e **III:** AA1C636_23 in azul e Mos8Chu0 em amarelo, with TM-score igual a 0,80812.

8.2.3 Análises Evolutivas

Avaliando a história evolutiva das glicoproteínas de Chuvirus endogenizados (Figura 14), foi possível perceber uma possível origem monofilética para este grupo de EVEs, uma vez que as formas circulantes atuais de vírus da família *Chuviridae* formam um clado próprio (sequências em amarelo). Outro ponto que deve ser observado é um agrupamento de EVEs de *Aedes aegypti* com um Chuvirus atual (Mos8Chu).

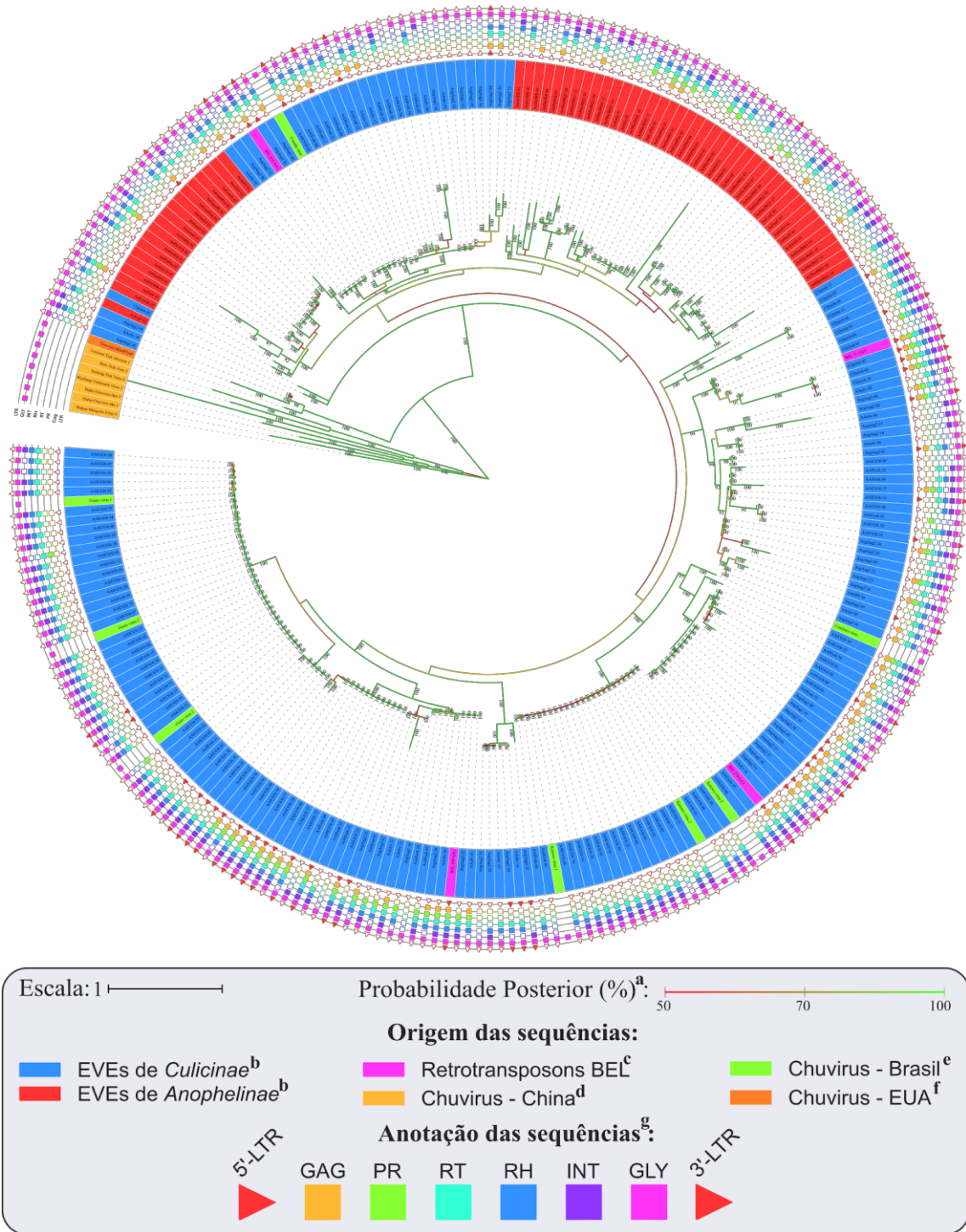
Analisando os clados de EVEs (sequências em azul e em vermelho), foi possível identificar 3 pontos principais:

- a) A presença de diferentes cópias de um mesmo elemento, indicado pelos clados com diversas sequências com tamanho do ramo em escala próxima a zero (o que indica a inexistência ou um valor próximo a 0 de substituições de aminoácidos);
- b) A presença de Retroelementos BEL-Pao (sequências em roxo) dentro e fora destes clados de cópias de EVEs;
- c) A presença de sequências descritas como *Chuvirus* atuais circulantes (sequências em verde) dentro e fora destes clados de cópias de EVEs.

Quando avaliamos a história evolutiva das regiões de transcriptase reversa e RNA-H de Retroelementos BEL-Pao com e sem glicoproteína de Chuvirus (Figura 15), outros 3 pontos principais ficam evidentes:

- a) A divergência entre essas duas categorias de elementos evidenciada pela formação de 2 clados diferentes (triângulos nos clados BEL e Pao);
- b) A definição dos retroelementos fusionados como pertencentes à família Pao, uma vez que todos os elementos que apresentaram glicoproteínas de Chuvirus em sua estrutura formaram um clado dentro do grande clado dos Pao (Figura 15, ramo 2);
- c) Uma maior quantidade de espécies que possuem Retroelementos com glicoproteína (16 espécies), quando comparados com Retroelementos sem glicoproteínas (9 espécies) (Figura 16).


Figura 14 - Representação filogenética de glicoproteínas de chuvirus endogenizados

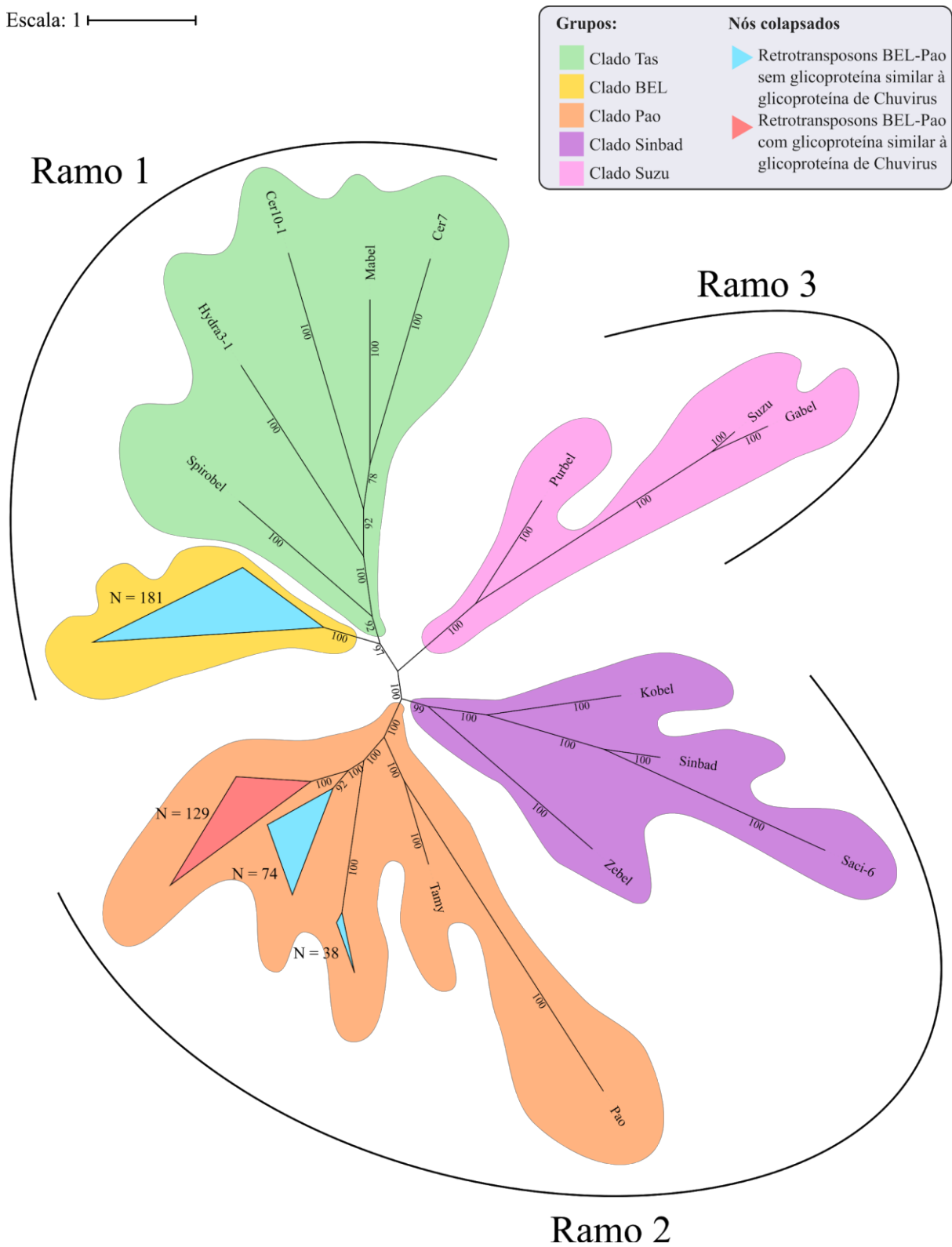


Fonte: O autor.

Legenda: Árvore bayesiana construída com 5.000.000 gerações a partir de 3 árvores iniciais com média de desvio padrão de ramo entre as árvores finais igual a 0,02. **a:** Probabilidade posterior em porcentagem, apenas valores maiores que 90 foram plotados na filogenia; **b:** Sequências identificadas como EVEs, neste trabalho, derivados da família viral Chuviridae; **c:** Sequências de retrotransposons da superfamília BEL presentes no RepBase e que apresentaram similaridade com glicoproteínas de Chuvirus; **d:** Primeiros chuvirus identificados na China por Li, C. et al 2015; **e:** Sequências descritas como Chuvirus no Brasil, por Pinto, A. et al 2017; **f:** Chuvirus depositado no NCBI sob código de acesso KX924631.1; **g:** LTR = *Long Terminal Repeats*; Gag = Domínio de proteínas de capsídeo; PR = Domínio de Protease; RT = Domínio de Transcriptase Reversa; RH = Domínio de Ribonuclease H; INT = Domínio de Integrase; GLY = Glicoproteína. Filogenia disponibilizada no link: https://itol.embl.de/tree/2001332_61329821563538693.

Figura 15 - Representação filogenética das regiões de Transcriptase Reversa e RNase H de poliproteínas de BEL-Pao.

Escala: 1 

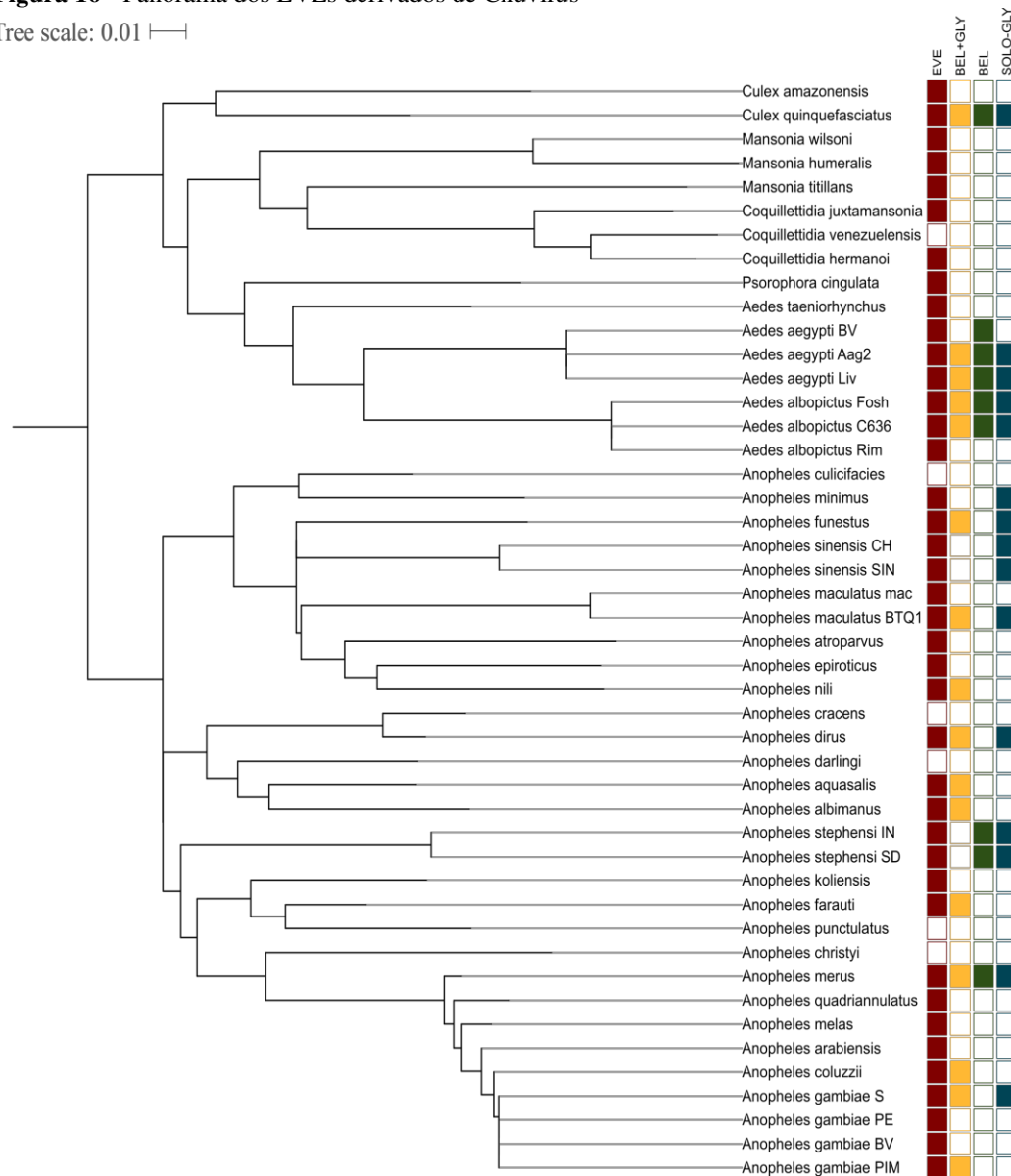


Fonte: O autor.

Legenda: Árvore bayesiana construída com 1.000.000 gerações a partir de 3 árvores iniciais com média de desvio padrão de ramo entre as árvores finais igual a 0,03. Ramos e clados foram anotados de acordo com os padrões de filogenia de retroelementos BEL-Pao. Filogenia completa disponibilizada no link <https://itol.embl.de/tree/200133261405581567779232>.

Figura 16 - Panorama dos EVEs derivados de Chuvirus

Tree scale: 0.01



Fonte: O autor.

Legenda: Blocos preenchidos representam presença enquanto blocos vazios representam ausência dos seguintes elementos: EVE, Retroelemento Pao com glicoproteína, Retroelemento BEL-Pao sem glicoproteína e apenas glicoproteína. Distribuição de espécies definidas por uma Análise filogenética de inferência bayesiana utilizando regiões nucleotídicas do gene Citocromo Oxidase I, com 1 milhão de gerações a partir de 3 árvores iniciais.

8.2.4 Validações das solo-glicoproteínas em bancada:

Foram avaliadas as regiões flanqueadoras de 5 elementos (**Tabela 4**), onde buscou-se identificar as regiões de ligação entre o EVE e o genoma hospedeiro. Visando avaliar a presença destes elementos em diferentes populações da mesma espécie, indivíduos das espécies *A. aegypti*, *A. albopictus* e *C. quinquefasciatus* tanto de insetário como de campo (Tabela 4) foram utilizados, bem como uma linhagem celular C6/36 de *A. albopictus*.

Tabela 4 - Elementos validados por PCR

Amostra	Local	Primer	Elemento*	Amplicon*	Identidade(%)**
AaeIns	IAM/Fiocruz	AAeliv_06_L	563	-	-
Aae1471	8°02'51.9"S 34°56'45.6"W	AAeliv_06_L	563	-	-
AaeIns	IAM/Fiocruz	AAeliv_06_R	729	629	99,68
Aae1471	8°02'51.9"S 34°56'45.6"W	AAeliv_06_R	729	634	99,53
AalZoo	8°00'43.3"S 34°56'40.7"W	AAIFosh_15_L	686	577	98,07
AalJB	8°04'36.9"S 34°57'34.1"W	AAIFosh_15_L	686	592	99,83
AalIns	IAM/Fiocruz	AAIFosh_15_L	686	585	93,04
AalC636	IAM/Fiocruz	AAIFosh_15_L	686	433	97,38
AalZoo	8°00'43.3"S 34°56'40.7"W	AAIFosh_15_R	771	-	-
AalJB	8°04'36.9"S 34°57'34.1"W	AAIFosh_15_R	771	-	-
AalIns	IAM/Fiocruz	AAIFosh_15_R	771	-	-
AalC636	IAM/Fiocruz	AAIFosh_15_R	771	-	-
AalZoo	8°00'43.3"S 34°56'40.7"W	AAIFosh_19_L	551	433	91,46
AalJB	8°04'36.9"S 34°57'34.1"W	AAIFosh_19_L	551	529	96,74
AalIns	IAM/Fiocruz	AAIFosh_19_L	551	498	92,53
AalC636	IAM/Fiocruz	AAIFosh_19_L	551	-	-
AalZoo	8°00'43.3"S 34°56'40.7"W	AAIFosh_19_R	672	-	-
AalJB	8°04'36.9"S 34°57'34.1"W	AAIFosh_19_R	672	-	-
AalIns	IAM/Fiocruz	AAIFosh_19_R	672	-	-
AalC636	IAM/Fiocruz	AAIFosh_19_R	672	-	-
AaeIns	IAM/Fiocruz	AegAag2_12_L	607	548	76,51
Aae1471	8°02'51.9"S 34°56'45.6"W	AegAag2_12_L	607	544	89,93
AaeIns	IAM/Fiocruz	AegAag2_12_R	800	696	90,97
Aae1471	8°02'51.9"S 34°56'45.6"W	AegAag2_12_R	800	515	88,15
Cxqui1304	8°02'51.9"S 34°56'45.6"W	CQuJoh_01_L	729	-	-
Cxqui1331	8°02'51.9"S 34°56'45.6"W	CQuJoh_01_L	729	-	-
CxIns	IAM/Fiocruz	CQuJoh_01_L	729	-	-
Cxqui1304	8°02'51.9"S 34°56'45.6"W	CQuJoh_01_R	695	669	92,96
Cxqui1331	8°02'51.9"S 34°56'45.6"W	CQuJoh_01_R	695	665	91,97
CxIns	IAM/Fiocruz	CQuJoh_01_R	695	667	97,49

Fonte: O autor.

Nota: * Tamanho em pb; ** Identidade do amplicon com o EVE correspondente.

9 DISCUSSÃO

Estudos anteriores que buscaram identificar EVEs em genomas de culicídeos (PALATINI *et al.*, 2017; RUSSO *et al.*, 2019; WHITFIELD *et al.*, 2017) utilizaram apenas vírus de RNA como sondas nas buscas por similaridade, porém, nossos resultados demonstram que a diversidade de EVEs presentes no genoma desses organismos abrange quase todas as classes de genoma viral (Figura 8), o que engloba vírus de DNA. Essa diversidade encontrada ressalta a necessidade de incluir o maior número possível de grupos virais, independente do seu tipo de material genético, nas análises de busca por EVEs.

Outro ponto relacionado à diversidade de EVEs encontrados, é a variedade de hospedeiros relacionados aos taxa virais identificados (**Figura 8D e E**), onde apesar de mosquitos normalmente estarem infectados e transmitir vírus de RNA (LI, CI-XIU *et al.*, 2015) que também infectam mamíferos e aves, foram identificados EVEs pertencentes a grupos virais que normalmente infectam plantas, fungos e procariotos, sugerindo que os processos de integração viral podem ser extremamente dinâmicos, não sendo necessária uma relação de longo prazo entre vírus e hospedeiro. Porém, apesar de todos os filtros utilizados neste estudo, análises mais minuciosas podem esclarecer se determinada sequência assinada como um EVE de procarioto em um eucarioto é realmente resultado de endogenização viral, ou apenas uma sequência que convergiu evolutivamente para uma sequência similar a um vírus.

Avaliando o baixo número de elementos identificados nos *datasets* privados gerados por sequenciamento de baixa cobertura (Figuras 9 e 10), foi possível avaliar que diferente de estudos de transposons (GOUBERT *et al.*, 2015), estratégias com dados de sequenciamento de baixa cobertura não são indicadas para estudos de EVEs, uma vez que apesar de que esses elementos possam estar relacionados a *transposons*, aparentemente isto não resulta num enriquecimento de sequências advindas de EVEs em sequenciamentos de baixa cobertura.

Um fator interessante que deve ser considerado ao estudar EVEs é a qualidade da montagem, uma vez que ao compararmos o conteúdo de EVEs de montagens mais robustas (que usam diferentes estratégias de sequenciamento, incluindo *long reads*), como as montagens Aag2 e C6/36 de *A. aegypti* e *A. albopictus*, respectivamente. Este ponto havia sido levantado por Whitfield e seus colaboradores (WHITFIELD *et al.*, 2017). Uma explicação plausível para esta diferença está na própria natureza dos EVEs, que podem ser integrados pela maquinaria de transposons ativos (HOLMES, 2009), o que resulta na presença desses EVEs em regiões repetitivas dos genomas enriquecidas por transposons

(RUSSO *et al.*, 2019; WHITFIELD *et al.*, 2017). Uma vez que regiões repetitivas são difíceis de montar utilizando sequenciamento de *short reads*, quantificações de EVEs em genomas com esse tipo de abordagem de sequenciamento (*short read*), podem gerar dados sub representados.

A identificação de agrupamentos sem diferença em tamanho de ramo na filogenia (Figura 14) das glicoproteínas entre EVEs e sequências supostamente de vírus circulantes indica que possivelmente Lara-Pinto e seus colaboradores (LARA PINTO *et al.*, 2017) identificaram na verdade material genético proveniente de EVEs como vírus circulantes, uma vez que o estudo foi realizado por meio de sequenciamento de RNA total de amostras de mosquitos, e não houve análise de cruzamento de dados com vírus endógenos, etapa que vem sendo considerada essencial em estudos de metavirômica para evitar a identificação de falsos positivos (NOURI *et al.*, 2018). Além da relação próxima destes vírus identificados por Lara-Pinto e seus colaboradores, da ausência da etapa de filtro com vírus endógenos, outro ponto que reforça a hipótese destas sequências serem EVEs é apenas a identificação de fragmentos, ou regiões truncadas, destes vírus (LARA PINTO *et al.*, 2017) e a alta identidade (mais de 99%, Apêndice C, Tabela S7) destes possíveis Chuvirus circulantes, com os EVEs identificados no presente estudo.

Levando em consideração as estruturas tridimensionais das proteínas modeladas, foi possível avaliar a similaridade das estruturas tridimensionais entre as glicoproteínas modeladas, o que pode levantar hipóteses da possibilidade destas glicoproteínas endogenizadas, estarem atuando no *fitness* do mosquito, uma vez que pode estar participando de um processo de imunidade viral mediada por competição por receptores, onde já foram relatados este tipo de processo, nos quais uma proteína de envelope viral de vírus endógenos compete com proteínas de vírus circulantes por receptores de membrana, em caprinos (*Ovis aries*), galinhas (*Gallus gallus domesticus*) e gatos (*Felis catus*) (ARMEZZANI *et al.*, 2014; ITO *et al.*, 2013; ROBINSON *et al.*, 1981).

Avaliando os dados que relacionam glicoproteínas hibridizadas a Retroelementos Pao, nossos resultados se enquadram nos dois modelos propostos de estrutura para elementos da superfamília BEL-Pao. O primeiro modelo, mais comumente encontrado, foi identificado em 1993 por Xiong e seus colaboradores (XIONG; BURKE; EICKBUSH, 1993) e não possui envelope viral em sua estrutura. O segundo modelo, que apresenta uma região codificante homóloga à um envelope viral foi proposto com base em três elementos, o elemento Tas (FELDER *et al.*, 1994) identificado em *Ascaris lumbricoides* com uma glicoproteína *Env-like* similar ao envelope viral de *Phlebovirus*, o elemento Cer-7 (BROWNING *et al.*, 1996)

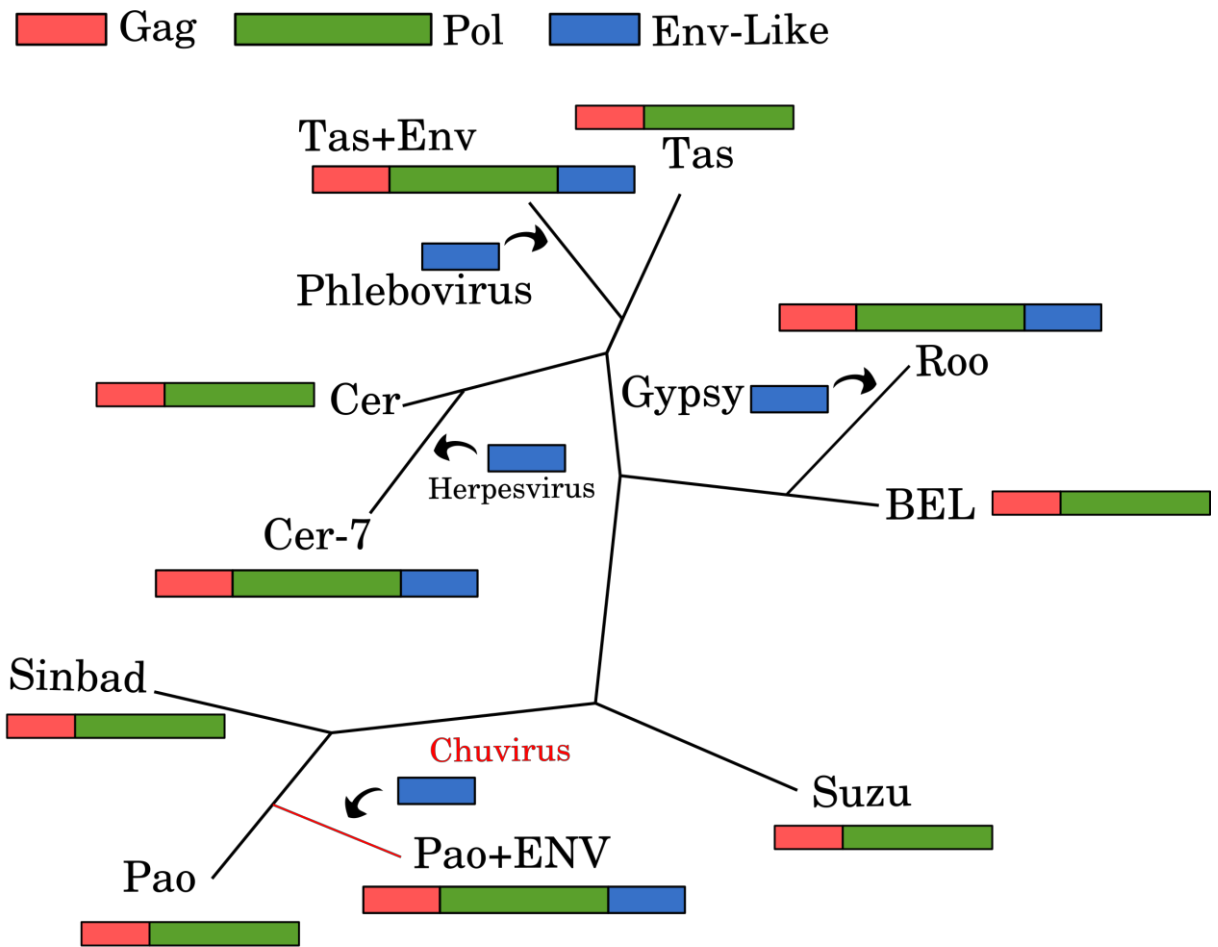
identificado em *Caenorhabditis elegans* com uma glicoproteína Env-like similar ao envelope de *Herpesvirus* e o elemento Roo (BEL-like) que possui uma glicoproteína similar à glicoproteínas identificadas em transposons da superfamília Gypsy (MALIK, HARMIT S.; HENIKOFF, 2005). Ambos os elementos parecem ter capturado estruturas de envelope viral em eventos independentes (MALIK, H. S.; HENIKOFF; EICKBUSH, 2000).

Outra análise que suporta a hipótese de que as várias cópias de glicoproteína encontrados nos genomas de culicídeos são provenientes da replicação dos elementos Pao pode ser observada na similaridade encontrada entre as LTRs das cópias potencialmente autônomas encontradas. Nossos resultados demonstram uma variação de similaridade entre as regiões LTRs entre 0,97 e 1 (Apêndice C, Tabela S5), o que pode ser explicado por eventos recentes de transposição. Após o momento de transposição, as LTR que até então eram idênticas, passam a acumular mutações independentemente, ou seja, a alta similaridade entre LTRs de uma cópia indicam a recente integração da mesma (DANGEL *et al.*, 1994). Assim, todas as informações obtidas nesse estudo corroboram que o alto número de sequências de glicoproteína de Chuvirus nos genomas dos mosquitos se deve a transposição de elementos Pao.

Em 2017 Whitfield e seus colaboradores (PALATINI *et al.*, 2017; WHITFIELD *et al.*, 2017) identificaram glicoproteínas de Chuvirus em genoma de *A. aegypti* flanqueadas por elementos da superfamília BEL e sugeriram que os elementos BEL do genoma de *A. aegypti* podem ter atuado como vetores de integração das glicoproteínas virais. No entanto, nossos resultados demonstram as glicoproteínas de Chuvirus como parte da própria estrutura de alguns os elementos da superfamília BEL-Pao, estando à jusante da poliproteína e entre as estruturas LTRs (Figura 11C e Apêndice C).

Nossos resultados, indicam a ocorrência de um quarto evento de recombinação entre elementos Pao e glicoproteínas de envelope viral (Figura 17), sendo o primeiro evento descrito em elementos Pao e em elementos encontrados no genoma de diferentes espécies de mosquitos. Até o momento, não foi possível encontrar elementos Pao associados a glicoproteínas de Chuvirus em outros genomas de eucariotos com exceção de culicídeos sugerindo que essa recombinação provavelmente aconteceu entre Chuvirus que infectam ou infectaram mosquitos e parasitas genômicos da superfamília BEL-Pao que estão presentes no genoma dessas espécies. Nossos achados reforçam que eventos de recombinação entre diferentes parasitas, mesmo parasitas evolutivamente distantes como vírus e elementos de transposição, podem dar origem a novos elementos funcionais capazes de se replicar no genoma hospedeiro.

Figura 17 - Panorama evolutivo atual de elementos da superfamília BEL/Pao.



Fonte: O autor baseado em Llorens *et al.* (2009).

Legenda: Ramo em vermelho representa o evento identificado neste estudo. Figura meramente ilustrativa.

10 CONCLUSÃO

Neste trabalho nós elucidamos a diversidade de EVEs nos genomas de culicídeos publicados até o momento. Nossos resultados demonstram que além desses EVEs advirem de uma diversidade viral não abordada em estudos anteriores, alguns grupos, como o caso dos Chuvirus, podem estar sendo tratados erroneamente como vírus circulantes em alguns estudos. Considerando os resultados relacionados a EVEs da família *Chuviridae*, foi possível verificar que estes elementos estão amplamente distribuídos pela família *Culicidae* e estão envolvidos em dois possíveis processos: a replicação de retroelementos Pao, através da aquisição de glicoproteínas de Chuvirus, e em uma possível resposta antiviral contra Chuvirus circulantes e retrovírus Pao, originados da hibridização entre retroelementos Pao e o gene de envelope de Chuvirus, devido a presença de solo-glicoproteínas em alguns dos genomas analisados e da similaridade das estruturas tridimensionais de glicoproteínas endogenizadas com as glicoproteínas presentes em Chuvirus circulantes e retrovírus Pao. Estes resultados trouxeram novos *insights* sobre a dinâmica evolutiva de EVEs e retrotransposons nos genomas de mosquitos e evidenciam a necessidade de novos estudos sobre o possível papel antiviral destes elementos à nível de proteína contra vírus circulantes.

11 CRONOGRAMA DE EXECUÇÃO

O presente trabalho seguiu o cronograma descrito no Quadro 1.

Quadro 1 – Cronograma do estudo

Atividades	2018			2019		
	Jan/Fev/ Mar/Abr	Mai/Jun/ Jul/Ago	Set/Out/ Nov/Dez	Jan/Fev/ Mar/Abr	Mai/Jun/ Jul/Ago	Set/Out/ Nov/Dez
Implementação de <i>Softwares</i>	x					
Busca por homologia nos genomas	x	x				
Extração de regiões flanqueadoras		x				
Desenho dos primers		x				
PCRs de validação		x	x			
Análises evolutivas				x	x	x
Contexto Genômico e Funcional				x	x	x
Escrita do manuscrito e defesa da dissertação			x	x	x	x

Fonte: O autor.

REFERÊNCIAS

- ABASCAL, F.; ZARDOYA, R.; TELFORD, M. J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. **Nucleic acids research**, Oxford, v. 38, p. W7-13, 30 jul. 2010.
- AIEWSAKUN, P.; KATZOURAKIS, A. Endogenous viruses: Connecting recent and ancient viral evolution. **Virology**, Washington, v. 479-480, p. 26-37, 12 maio 2015.
- AJAWATANAWONG, P. Molecular Phylogenetics: Concepts for a Newcomer. **Advances in biochemical engineering/biotechnology**, Paris, v. 160, p. 185-196, 2016.
- ALLEN, A. *et al.* Evolution of glutamate dehydrogenase regulation of insulin homeostasis is an example of molecular exaptation. **Biochemistry**, Berkeley, v. 43, n. 45, p. 14431-14443, 16 nov. 2004.
- ALTSCHUL, S. F. *et al.* Basic Local Alignment Search Tool. **Journal of Molecular Biology**, Amsterdam, v. 215, p. 403-410, 1988.
- ANDREWS, S. FastQC A Quality Control tool for High Throughput Sequence Data. 2011. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Acesso em: 15 out. 2018.
- ARCHIBALD, J. K.; MORT, M. E.; CRAWFORD, D. J. Bayesian Inference of Phylogeny: A Non-Technical Primer. **Taxon**, Nova Iorque, v. 52, n. 2, 1 maio 2003.
- ARMEZZANI, A. *et al.* “Ménage à Trois”: the evolutionary interplay between JSRV, enJSRVs and domestic sheep. **Viruses**, Basiléia, v. 6, n. 12, p. 4926-4945, 9 dez. 2014.
- ASAD, H.; CARPENTER, D. O. Effects of climate change on the spread of zika virus: a public health threat. **Reviews on environmental health**, Berlin, v. 33, n. 1, p. 31-42, 28 mar. 2018.
- ASWAD, A.; KATZOURAKIS, A. Paleovirology: The Study of Endogenous Viral Elements. **Virus Evolution: Current Research and Future Directions**: Caister Academic, Oxford, 2015. p. 273-292.
- ATONI, E. *et al.* Metagenomic Virome Analysis of Culex Mosquitoes from Kenya and China. **Viruses**, Basiléia, v. 10, n. 1, 12 jan. 2018.
- BAHN, Peter R. The Tangled Tree: a Radical New History of Life by David Quammen, Simon & Schuster, 2018. **Origins of life and evolution of the biosphere**: the journal of the International Society for the Study of the Origin of Life, Zwijndrecht, v. 48, n. 3, p. 343-344, 27 set. 2018.
- BALDAUF, S.L. Phylogeny for the faint of heart: a tutorial. **Trends in Genetics**, Cambridge, v. 19, n. 6, p. 345-351, 2002
- BALTIMORE, D. Expression of animal virus genomes. **Bacteriological reviews**, Washington, v. 35, n. 3, p. 235-41, 1970.
- BANKEVICH, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to

single-cell sequencing. **Journal of computational biology: a journal of computational molecular cell biology**, Londres, v. 19, n. 5, p. 455-477, 16 maio 2012.

BAQUERIZO-AUDIOT, E. *et al.* Structure and expression strategy of the genome of *Culex pipiens* densovirus, a mosquito densovirus with an ambisense organization. **Journal of virology**, Irvine, v. 83, n. 13, p. 6863-6873, 22 jul. 2009.

BARRETT, A. D. T. The reemergence of yellow fever. **Science**, Nova Iorque, v. 361, n. 6405, p. 847-848, 31 ago. 2018.

BASAK, S. *et al.* Reviewing Chandipura: a vesiculovirus in human epidemics. **Bioscience reports**, Londres, v. 27, n. 4-5, p. 275-298, 1 out. 2007.

BEER, B.; KURTH, R.; BUKREYEV, A. Characteristics of Filoviridae: Marburg and Ebola viruses. **Die Naturwissenschaften**, Basingstoke, v. 86, n. 1, p. 8-17, 1 jan. 1999.

BESANSKY, N. J. *et al.* Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 100, n. 19, p. 10818-10823, 16 set. 2003.

BLAIR, C. D.; OLSON, K. E. The role of RNA interference (RNAi) in arbovirus-vector interactions. **Viruses**, Basiléia, v. 7, n. 2, p. 820-843, 17 fev. 2015.

BLAISE, S. *et al.* Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 100, n. 22, p. 13013-13018, 28 out. 2003.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, Oxford, v. 30, n. 15, p. 2114-2120, 1 ago. 2014.

BRAGA, I. A. *et al.* *Aedes aegypti* resistance to temephos during 2001 in several municipalities in the states of Rio de Janeiro, Sergipe, and Alagoas, Brazil. **Memorias do Instituto Oswaldo Cruz**, Rio de Janeiro, v. 99, n. 2, p. 199-203, 1 mar. 2004.

BRONKHORST, A. W.; RIJ, R. P. VAN. The long and short of antiviral defense: small RNA-based immunity in insects. **Current opinion in virology**, Amsterdã, v. 7, p. 19-28, 15 ago. 2014.

BROWNING, H. *et al.* Macrorestriction analysis of *Caenorhabditis elegans* genomic DNA. **Genetics**, Maryland, v. 144, n. 2, p. 609-619, 1 out. 1996.

BUKHARI, K. *et al.* Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abysoviridae, and from a sister group to the Coronavirinae, the proposed genus Alphaletovirus. **Virology**, Washington, v. 524, p. 160-171, 7 nov. 2018.

BULIVA, E. *et al.* Emerging and Reemerging Diseases in the World Health Organization (WHO) Eastern Mediterranean Region-Progress, Challenges, and WHO Initiatives. **Frontiers in public health**, Lousane, v. 5, p. 276, 19 out. 2017.

- CAMINADE, C. *et al.* Impact of climate change on global malaria distribution. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 111, n. 9, p. 3286-3291, 4 mar. 2014.
- CARISSIMO, G. *et al.* Identification and Characterization of Two Novel RNA Viruses from Anopheles gambiae Species Complex Mosquitoes. **PloS one**, San Francisco, v. 11, n. 5, p. e0153881, 3 maio 2016.
- CASTRESANA, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. **Molecular Biology and Evolution**, Oxford, v. 17, p. 540-552. 2000.
- CHARLES, J. *et al.* Detection of novel and recognized RNA viruses in mosquitoes from the Yucatan Peninsula of Mexico using metagenomics and characterization of their in vitro host ranges. **The Journal of general virology**, Londres, v. 99, n. 12, p. 1729-1738, 9 dez. 2018.
- CHEN, R. *et al.* ICTV Virus Taxonomy Profile: Togaviridae. **The Journal of general virology**, Londres, v. 99, n. 6, p. 761-762, 10 jun. 2018.
- CHEN, X.-G. *et al.* Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 112, n. 44, p. E5907-E5915, 3 nov. 2015.
- CHOUIN-CARNEIRO, T. *et al.* Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. **PLoS neglected tropical diseases**, San Francisco, v. 10, n. 3, p. e0004543, 3 mar. 2016.
- CLEMENTS, A. N. The Biology of Mosquitoes. Wallingford, **Oxon**, Oxford, UK: CABI, 2000.
- CLETON, N. *et al.* Come fly with me: review of clinically important arboviruses for global travelers. **Journal of clinical virology: the official publication of the Pan American Society for Clinical Virology**, Berlin, v. 55, n. 3, p. 191-203, 25 nov. 2012.
- COFFEY, L. L. *et al.* Enhanced arbovirus surveillance with deep sequencing: Identification of novel rhabdoviruses and bunyaviruses in Australian mosquitoes. **Virology**, Washington, v. 448, p. 146-158, 5 jan. 2014.
- CONSOLI, R. A. G. B.; OLIVEIRA, R. L. Principais mosquitos de importância sanitária no Brasil. Rio de Janeiro: Fiocruz, 1994.
- CROCHU, S. *et al.* Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. **The Journal of general virology**, Londres, v. 85, n. Pt 7, p. 1971-1980, 1 jul. 2004.
- DANGEL, A. W. *et al.* Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution. **Immunogenetics**, Berlin, v. 42, n. 1, p. 41-52, 1994.
- DARWIN, C. R. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London: John Murray, 1859.

DAVISON, A. J. *et al.* The order Herpesvirales. **Archives of virology**, Berlin, v. 154, n. 1, p. 171-177, 9 dez. 2009.

DOTTO, B. R. *et al.* HTT-DB: new features and updates. **Database: the journal of biological databases and curation**, Oxford, v. 2018, 1 jan. 2018.

DUFFY, S.; SHACKELTON, L. A.; HOLMES, E. C. Rates of evolutionary change in viruses: patterns and determinants. **Nature Reviews Genetics**, Basingstoke, v. 9, n. 4, p. 267-276, 2007.

DUPRESSOIR, A. *et al.* Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 102, n. 3, p. 725-730, 18 jan. 2005.

EHRENFELD, E.; DOMINGO, E.; ROSS, R. The picornaviruses. Washington, DC: ASM Press, 2010.

EMERMAN, M.; MALIK, H. S. Paleovirology--modern consequences of ancient viruses. **PLoS biology**, San Francisco, v. 8, n. 2, p. e1000301, 9 fev. 2010.

EWELS, P. *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. **Bioinformatics**, Oxford, v. 32, n. 19, 1 out. 2016.

EYNGOR, M. *et al.* Identification of a novel RNA virus lethal to tilapia. **Journal of clinical microbiology**, Londres, v. 52, n. 12, p. 4137-4146, 17 dez. 2014.

FELDER, H. *et al.* Tas, a retrotransposon from the parasitic nematode *Ascaris lumbricoides*. **Gene**, Berlin, v. 149, n. 2, p. 219-225, 18 nov. 1994.

FESCHOTTE, C.; GILBERT, C. Endogenous viruses: insights into viral evolution and impact on host biology. **Nature reviews. Genetics**, Basingstoke, v. 13, n. 4, p. 283-296, 16 mar. 2012.

FINLAY, C. The Mosquito Hypothetically Considered as an Agent in the Transmission of Yellow Fever Poison. **The Yale journal of biology and medicine**, Connecticut, v. 9, n. 6, p. 589-604, 1 jul. 1937.

FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: interactive sequence similarity searching. **Nucleic acids research**, Oxford, v. 39, n. Web Server issue, p. W29-W37, 18 jul. 2011.

FORATTINI, O P. *Culicidologia Médica*. São Paulo: **Edusp**, 2002.

FORATTINI, O. P. *Ecologia, epidemiologia e sociedade*. São Paulo: **Artes Médicas**, 1992.

FORT, P. *et al.* Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. **Molecular biology and evolution**, Oxford, v. 29, n. 1, p. 381-390, 13 jan. 2012.

FRANK, J. A.; FESCHOTTE, C. Co-option of endogenous viral sequences for host cell function. **Current opinion in virology**, Amsterdã, v. 25, p. 81-89, 16 ago. 2017.

FREY, K. G. *et al.* Bioinformatic Characterization of Mosquito Viromes within the Eastern United States and Puerto Rico: Discovery of Novel Viruses. **Evolutionary bioinformatics online**, Los Angeles, v. 12, n. Suppl 2, p. 1-12, 16 jun. 2016.

GEL, B.; SERRA, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. **Bioinformatics**, Oxford, v. 33, n. 19, p. 3088-3090, 1 out. 2017.

GORBALENYA, A. E. *et al.* Nidovirales: evolving the largest RNA virus genome. **Virus research**, Berlin, v. 117, n. 1, p. 17-37, 28 abr. 2006.

GOUBERT, C. *et al.* De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). **Genome biology and evolution**, Oxford, v. 7, n. 4, p. 1192-1205, 11 mar. 2015.

GOULD, S. J.; VRBA, E. S. Exaptation-A missing term in the science of form. **Paleobiology**, Cambridge, v. 8, p. 4-15, 1981.

GRARD, G. *et al.* Genomics and evolution of Aedes-borne flaviviruses. **The Journal of general virology**, Londres, v. 91, n. Pt 1, p. 87-94, 9 jan. 2010.

GRASSI, B. **Studi di uno zoologo sulla malária**. [S. l.]: R. Accademia dei lincei, 1900.

GUARDADO-CALVO, P.; REY, F. A. The Envelope Proteins of the Bunyavirales. **Advances in virus research**, Berlin, v. 98, p. 83-118, 8 abr. 2017.

GUBLER, D. J. The changing epidemiology of yellow fever and dengue, 1900 to 2003: full circle? **Comparative immunology, microbiology and infectious diseases**, Berlin, v. 27, n. 5, p. 319-330, 1 set. 2004.

GULERIA, A. *et al.* Reviewing host proteins of Rhabdoviridae: possible leads for lesser studied viruses. **Journal of biosciences**, Bangalor, v. 36, n. 5, p. 929-937, 1 dez. 2011.

HAIG, D. A.; WOODALL, J. P.; DANSKIN, D. Thogoto virus: a hitherto underscribed agent isolated from ticks in kenya. **Journal of general microbiology**, Londres, v. 38, p. 389-394, 1 mar. 1965.

HANCOCK, J. M. Primer3. *In*: HANCOCK, J. M.; ZVELEBIL, M. J. (ed.). **Dictionary of Bioinformatics and Computational Biology**. John Wiley & Sons, 2004.

HANCOCK, P. A. *et al.* Associated patterns of insecticide resistance in field populations of malaria vectors across Africa. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 115, n. 23, p. 5938-5943, 5 jun. 2018.

HOLMES, E. C. The evolution of endogenous viral elements. **Cell host & microbe**, Maryland, v. 10, p. 368-377, 2009.

INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES. Reorganization and expansion of the order Nidovirales at the family and sub-order ranks. **ICTV**, Londres, 2017. Disponível em: <https://talk.ictvonline.org/>. Acesso em: 15 jan. 2018.

ITO, J. *et al.* Refrex-1, a soluble restriction factor against feline endogenous and exogenous retroviruses. **Journal of virology**, Irvine, v. 87, n. 22, p. 12029-12040, 21 nov. 2013.

JOLY-LOPEZ, Z.; BUREAU, T. E. Exaptation of transposable element coding sequences. **Current opinion in genetics & development**, Berlin, v. 49, p. 34-42, 8 abr. 2018.

JURKA, J. Repbase update: a database and an electronic journal of repetitive elements. **Trends in Genetics**, Cambridge, v. 16, n. 9, p. 418-20, 1999.

KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nature Reviews Genetics**. Basingstoke, v. 9, p. 411-412, 2008.

KATOH, K. *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, Oxford, v. 30, n. 14, p. 3059-3066, 2001.

KATZOURAKIS, A. Editorial overview: Paleovirology: the genomic fossil record, and consequences of ancient viral infections. **Current opinion in virology**, Amsterdã, v. 25, p. ix-xi, 1 ago. 2017.

KATZOURAKIS, A.; GIFFORD, R. J. Endogenous viral elements in animal genomes. **PLoS genetics**, San Francisco, v. 6, n. 11, p. e1001191, 18 nov. 2010.

KEARSE, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, Oxford, v. 28, n. 12, p. 1647-1649, 15 jun. 2012.

KELLEY, L. A. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. **Nature protocols**, Basingstoke, v. 10, n. 6, p. 845-858, 7 jun. 2015.

KIDWELL, M. G. M. G.; LISCH, D. R. D. R. Perspective: transposable elements, parasitic DNA, and genome evolution. **Evolution**, Hoboken, v. 55, n. 1, p. 1-24, 2000.

KIMURA, M.; OHTA, T. On some principles governing molecular evolution. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 71, n. 7, p. 2848-2852, 1 jul. 1974.

KINSLEY, A. C. *et al.* Mosquitoes in Moose Country: A Mosquito Survey of Northern Minnesota. **Journal of the American Mosquito Control Association**, San Bernardino, v. 32, n. 2, p. 83-90, 1 jun. 2016.

KNIPE, D. M.; HOWLEY, P. M. **Fields virology**. 6^a Ed., LLW, 2582p. 2013.

KOONIN, E. V.; DOLJA, V. V. Virus world as an evolutionary network of viruses and capsidless selfish elements. **Microbiology and molecular biology reviews: MMBR**, Berlin, v. 78, n. 2, p. 278-303, 1 jun. 2014.

KOONIN, E. V.; DOLJA, V. V. Metaviromics: a tectonic shift in understanding virus evolution. **Virus research**, Berlin, v. 246, p. A1-A3, 15 fev. 2018.

KOONIN, E. V.; DOLJA, V. V.; KRUPOVIC, M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. **Virology**, Washington, v. 479-480, p. 2-25, 12 maio

2015.

KRUPOVIC, M. *et al.* Ortervirales: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses. **Journal of virology**, Irvine, v. 92, n. 12, 15 jun. 2018.

LAFFERTY, K. D.; MORDECAI, E. A. The rise and fall of infectious disease in a warmer world. **F1000Research**, Londres, v. 5, 19 ago. 2016.

LAMPRECHT, B. *et al.* Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. **Nature medicine**, Basingstoke, v. 16, n. 5, p. 571-9, 1p following 579, 2 maio 2010.

LARA PINTO, A. Z. DE *et al.* Novel viruses in salivary glands of mosquitoes from sylvatic Cerrado, Midwestern Brazil. **PloS one**, San Francisco, v. 12, n. 11, p. e0187429, 8 nov. 2017.

LARSSON, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. **Bioinformatics**, Oxford, v. 30, n. 22, 15 nov. 2014.

LEACHE, A. D. The Timetree of Life. S. Blair Hedges and Sudhir Kumar, editors. **Integrative and Comparative Biology**, Oxford, v. 50, n. 1, 1 jul. 2010.

LEFORT, V.; LONGUEVILLE, J.-E.; GASCUEL, O. SMS: Smart Model Selection in PhyML. **Molecular biology and evolution**, Oxford, v. 34, n. 9, p. 2422-2424, 1 set. 2017.

LEQUIME, S.; LAMBRECHTS, L. Discovery of flavivirus-derived endogenous viral elements in mosquito genomes supports the existence of -associated insect-specific flaviviruses. **Virus evolution**, Oxford, v. 3, n. 1, p. vew035, 5 jan. 2017.

LEQUIME, S.; PAUL, R. E.; LAMBRECHTS, L. Determinants of Arbovirus Vertical Transmission in Mosquitoes. **PLoS pathogens**, San, Francisco, v. 12, n. 5, p. e1005548, 12 maio 2016.

LETUNIC, I.; BORK, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. **Bioinformatics**, Oxford, v. 23, n. 1, p. 127-128, 1 jan. 2007.

LI, C.-X. *et al.* Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. **eLife**, Cambridge, v. 4, 29 jan. 2015.

LLORENS, C. *et al.* The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. **Nucleic acids research**, Oxford, v. 39, n. Database issue, p. D70-D74, 29 jan. 2011.

LOEW, H.; MEIGEN, J. W. **Systematische Beschreibung der bekannten europäischen zweiflügeligen Insekten**. F. W. Forstmann, 1817.

LUO, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. **GigaScience**, Oxford, v. 1, n. 1, p. 18, 27 dez. 2012.

MAES, P. *et al.* Taxonomy of the family Arenaviridae and the order Bunyavirales: update 2018. **Archives of virology**, Berlin, v. 163, n. 8, p. 2295-2310, 21 ago. 2018.

MALIK, H. S.; HENIKOFF, S. Positive selection of Iris, a retroviral envelope-derived host

gene in *Drosophila melanogaster*. **PLoS genetics**, San Francisco, v. 1, n. 4, p. e44, 1 out. 2005.

MALIK, H. S.; HENIKOFF, S.; EICKBUSH, T. H. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. **Genome research**, Oxford, v. 10, n. 9, p. 1307-1318, 1 set. 2000.

MARQUES, J. T. *et al.* Functional specialization of the small interfering RNA pathway in response to virus infection. **PLoS pathogens**, San Francisco, v. 9, n. 8, p. e1003579, 29 ago. 2013.

MASSAD, E.; MIGUEL, M. M.; COUTINHO, F. A. B. Is vaccinating monkeys against yellow fever the ultimate solution for the Brazilian recurrent epizootics? **Epidemiology and infection**, Cambridge, v. 146, n. 13, p. 1622-1624, 14 out. 2018.

MENDEL, G. Experiments in Plant Hybridization. **Scholarly Publishing**, n. 1865, 1864.

MI, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. **Nature**, Basingstoke, v. 403, n. 6771, p. 785-789, 17 fev. 2000.

MOSQUITO TAXONOMY INVENTORY. Valid Species. 2008. Disponível em: <http://mosquito-taxonomic-inventory.info/valid-species-list>. Acesso: 14 Dez. 2019.

NCBI, 2018. Genomes. Disponível em: <https://www.ncbi.nlm.nih.gov/genome>. Acesso em: 17 Ago. 2018.

NIBERT, M. L.; FURLONG, D. B.; FIELDS, B. N. Mechanisms of viral pathogenesis. Distinct forms of reoviruses and their roles during replication in cells and host. **The Journal of clinical investigation**, Ann Arbor, v. 88, n. 3, p. 727-734, 1 set. 1991.

NOURI, S. *et al.* Insect-specific viruses: from discovery to potential translational applications. **Current opinion in virology**, Amsterdã, v. 33, p. 33-41, 23 dez. 2018.

NOVÁK, P. *et al.* RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. **Bioinformatics**, Oxford, v. 26, p. 1-2, 2012.

OKONECHNIKOV, K.; GOLOSOVA, O.; FURSOV, M. Unipro UGENE: a unified bioinformatics toolkit. **Bioinformatics**, Oxford, v. 28, n. 8, 15 abr. 2012.

ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2018. Disease Outbreak by year. Disponível em: <https://www.who.int/csr/don/archive/year/en>. Acesso em 14 Set. 2018.

PALATINI, U. *et al.* Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors *Aedes aegypti* and *Aedes albopictus*. **BMC genomics**, Oxford, v. 18, n. 1, p. 512, 5 jul. 2017.

PARSEVAL, N. DE; HEIDMANN, T. Human endogenous retroviruses: from infectious elements to human genes. **Cytogenet Genome Res**, Basel, v. 110, p. 318-32, 1 jan. 2005.

PASTERNAK, A. O. Nidovirus transcription: how to make sense...? **Journal of General Virology**, Oxford, v. 87, n. 6, 1 jun. 2006.

- PATEL, M. R.; EMERMAN, M.; MALIK, H. S. Paleovirology—ghosts and gifts of viruses past. **Current opinion in virology**, Amsterdã, v. 1, p. 304-309, 2009.
- PFEIFFER, J. K.; KIRKEGAARD, K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. **PLoS pathogens**, San Francisco, v. 1, n. 2, p. e11, 7 out. 2005.
- PHILBEY, A. W. *et al.* An apparently new virus (family Paramyxoviridae) infectious for pigs, humans, and fruit bats. **Emerging infectious diseases**, Berlin, v. 4, n. 2, p. 269-271, 1 jun. 1998.
- PRESTI, R. M. *et al.* Quarantil, Johnston Atoll, and Lake Chad viruses are novel members of the family Orthomyxoviridae. **Journal of virology**, Irvine, v. 83, n. 22, p. 11599-11606, 2 nov. 2009.
- PRINGLE, C. R. The order Mononegavirales--current status. **Archives of virology**, Berlin, v. 142, n. 11, p. 2321-2326, 1996.
- QUITO-AVILA, D. F. *et al.* Genetic characterization of Blueberry necrotic ring blotch virus, a novel RNA virus with unique genetic features. **The Journal of general virology**, Londres, v. 94, n. Pt 6, p. 1426-1434, 13 jun. 2013.
- REN, X.; HOICZYK, E.; RASGON, J. L. Viral paratransgenesis in the malaria vector *Anopheles gambiae*. **PLoS pathogens**, San Francisco, v. 4, n. 8, p. e1000135, 22 ago. 2008.
- RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: the European Molecular Biology Open Software Suite. **Trends Genet**, Berlin, v. 16, p. 276-277, 1 jun. 2000.
- ROBINSON, H. L. *et al.* Host Susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. **Journal of virology**, Irvine, v. 40, n. 3, p. 745-751, 1 dez. 1981.
- RONQUIST, F.; HUELSENBECK, J.; TESLENKO, M. Draft Mr. Bayes version 3.2 Manual, 2010. Disponível em: http://mrbayes.sourceforge.net/mb3.2_manual.pdf. Acesso em: 18 ago. 2018.
- ROUNDY, C. M. *et al.* Insect-Specific Viruses: A Historical Overview and Recent Developments. **Advances in virus research**, Oxford, v. 98, p. 119-146, 17 nov. 2017.
- RUSSO, A. G. *et al.* Novel insights into endogenous RNA viral elements in and other arbovirus vector genomes. **Virus evolution**, Oxford, v. 5, n. 1, p. vez010, 18 jan. 2019.
- SANJUÁN, R. *et al.* Viral mutation rates. **Journal of virology**, Irvine, v. 84, n. 19, p. 9733-9748, 21 out. 2010.
- SCALLY, A. The mutation rate in human evolution and demographic inference. **Current opinion in genetics & development**, Berlin, v. 41, p. 36-43, 31 dez. 2016.
- SHAIKEVICH, E. V. *et al.* Genetic diversity of *Culex pipiens* mosquitoes in distinct populations from Europe: contribution of *Cx. quinquefasciatus* in Mediterranean populations. **Parasites & vectors**, Basingstok, v. 9, p. 47, 27 jan. 2016.

SHI, M. *et al.* Redefining the invertebrate RNA virosphere. **Nature**, Basingstoke, v. 540, n. 7634, p. 539-543, 22 dez. 2016.

SHI, M. *et al.* High-Resolution Metatranscriptomics Reveals the Ecological Dynamics of Mosquito-Associated RNA Viruses in Western Australia. **Journal of virology**, Irvine, v. 91, n. 17, 1 set. 2017.

SHI, M. *et al.* The evolutionary history of vertebrate RNA viruses. **Nature**, Basingstoke, v. 556, n. 7700, p. 197-202, 4 abr. 2018.

SILVA, A. Sequenciamento de genomas mitocondriais de mosquitos provenientes de áreas silvestres. Recife: 2019. 112 p.

SIMMONDS, P. *et al.* ICTV Virus Taxonomy Profile: Flaviviridae. **The Journal of general virology**, Londres, v. 98, n. 1, p. 2-3, 1 jan. 2017.

SOUCY, S. M.; HUANG, J.; GOGARTEN, J. P. Horizontal gene transfer: building the web of life. **Nature reviews. Genetics**, Basingstoke, v. 16, n. 8, p. 472-482, 1 ago. 2015.

STOLLAR, V.; THOMAS, V. L. An agent in the *Aedes aegypti* cell line (Peleg) which causes fusion of *Aedes albopictus* cells. **Virology**, Oxford, v. 64, n. 2, p. 367-377, 1 abr. 1975.

TAKKEN, W.; VERHULST, N. O. Host preferences of blood-feeding mosquitoes. **Annual review of entomology**, Cambridge, v. 58, p. 433-453, 27 set. 2013.

TASSETTO, M. *et al.* Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements. **eLife**, Cambridge, v. 8:e41244, p. 1-29, 2018.

TATUM, E. L.; LEDERBERG, J. Gene Recombination in the Bacterium *Escherichia coli*. **Journal of bacteriology**, Berlin, v. 53, n. 6, p. 673-684, 1 jun. 1947.

TEMIN, H. M. Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. **Molecular biology and evolution**, Oxford, v. 2, n. 6, p. 455-468, 1 nov. 1985.

TIZARD, I. *et al.* The pathogenesis of bornaviral diseases in mammals. **Animal health research reviews**, Cambridge, v. 17, n. 2, p. 92-109, 23 dez. 2016.

TOLLE, M. A. Mosquito-borne diseases. **Current problems in pediatric and adolescent health care**, Berlin, v. 39, n. 4, p. 97-140, 1 abr. 2009.

TSANGARAS, K.; GREENWOOD, A. D. Paleovirology: Viral Sequences from Historical and Ancient DNA. In: **Paleogenomics: Genome-Scale Analysis of Ancient DNA**, 2017.

TU, Z.; HILL, J. J. MosquI, a novel family of mosquito retrotransposons distantly related to the *Drosophila* I factors, may consist of elements of more than one origin. **Molecular biology and evolution**, Oxford, v. 16, n. 12, p. 1675-1686, 1 dez. 1999.

VALLES, S. M. *et al.* ICTV Virus Taxonomy Profile: Dicistroviridae. **The Journal of general virology**, Londres, v. 98, n. 3, p. 355-356, 1 mar. 2017.

VECTORBASE. 2018. Disponível em: <https://www.vectorbase.org>. Acesso em: 15 Jun. 2018.

VERLI, H (org.). *Bioinformática da Biologia à flexibilidade*. 1. Ed. - São Paulo: SBBq, 2014. 282 p.

VIGNUZZI, M. *et al.* Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. **Nature**, Basingstoke, v. 439, n. 7074, 4 dez. 2005.

WALLACE, A. R. On the law which has regulated the introduction of new species. **Annals and Magazine of Natural History**, Londres, v. 16, n. 93, 1 set. 1855.

WEAVER, S. C.; BARRETT, A. D. T. Transmission cycles, host range, evolution and emergence of arboviral disease. **Nature reviews. Microbiology**, Basingstoke, v. 2, n. 10, p. 789-801, 1 out. 2004.

WEBSTER, J. P.; BORLASE, A.; RUDGE, J. W. Who acquires infection from whom and how? Disentangling multi-host and multi-mode transmission dynamics in the “elimination” era. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, Londres, v. 372, n. 1719, 5 maio 2017.

WHITFIELD, Z. J.; DOLAN, P. T., KUNITOMI, M. *et al.* The Diversity, Structure, and Function of Heritable Adaptive Immunity Sequences in the *Aedes aegypti* Genome. **Current biology**, Oxford, v. 27, n. 22, p. 3511-3519.e7, 20 nov. 2017.

XIA, H. *et al.* Comparative Metagenomic Profiling of Viromes Associated with Four Common Mosquito Species in China. **Virologica Sinica**, Berlin, v. 33, n. 1, p. 59-66, 2 fev. 2018.

XIAO, P. *et al.* Metagenomic Sequencing From Mosquitoes in China Reveals a Variety of Insect and Human Viruses. **Frontiers in cellular and infection microbiology**, Oxford, v. 8, p. 364, 19 out. 2018.

XIONG, Y.; BURKE, W. D.; EICKBUSH, T. H. Pao, a highly divergent retrotransposable element from *Bombyx mori* containing long terminal repeats with tandem copies of the putative R region. **Nucleic acids research**, Oxford, v. 21, n. 9, p. 2117-2123, 11 maio 1993.

XU, Z.; WANG, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. **Nucleic acids research**, Oxford, v. 35, p. W265-W268, 7 jul. 2007.

YE, J.; COULOURIS, G.; ZARETSKAYA, I. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. **BMC bioinformatics**, Oxford, v. 13, p. 134, 2011.

KUHN, J. H.; WOLF, Y; KRUPOVIC, M. *et al.* **Megataxonomy of negative-sense RNA viruses**. ICTV, Londres. 2017.

ZERBINO, D. R.; BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. **Genome research**, Cambridge, v. 18, n. 5, p. 821-829, 18 maio 2008.

ZHANG, Y.; SKOLNICK, J. TM-Align: A Protein Structure Alignment Algorithm Based

on the TM-Score. **Nucleic Acids Research**, Oxford, v. 33, n. 7, p. 2302-2309, 1 jan. 2005.

APÊNDICE – INFORMAÇÕES ADICIONAIS

A. Genomas de referência para figuras 2 e 3.

Tabela suplementar 1: Genomas de representantes de grupos virais já relatados infectando mosquitos

Vírus	Referência
<i>Culex pipiens</i> Dengovirus	https://www.ncbi.nlm.nih.gov/nuccore/NC_012685
<i>Equine encephalosis virus</i>	https://www.ncbi.nlm.nih.gov/assembly/GCF_002829385.1
<i>Dengue virus</i>	https://www.ncbi.nlm.nih.gov/nuccore/NC_001474
<i>Anopheles C virus</i>	https://www.ncbi.nlm.nih.gov/nuccore/KU169878.1
<i>Houston virus</i>	https://www.ncbi.nlm.nih.gov/nuccore/MH443059.1
<i>Chikungunya virus</i>	https://www.ncbi.nlm.nih.gov/assembly/260971
<i>Wuhan Mosquito Vírus 8</i>	https://www.ncbi.nlm.nih.gov/assembly/GCF_001440835.1
<i>Arboretum almendravirus</i>	https://www.ncbi.nlm.nih.gov/assembly/GCF_000927335.1
<i>Thogoto virus</i>	https://www.ncbi.nlm.nih.gov/assembly/261071

B. Qualidade dos datasets



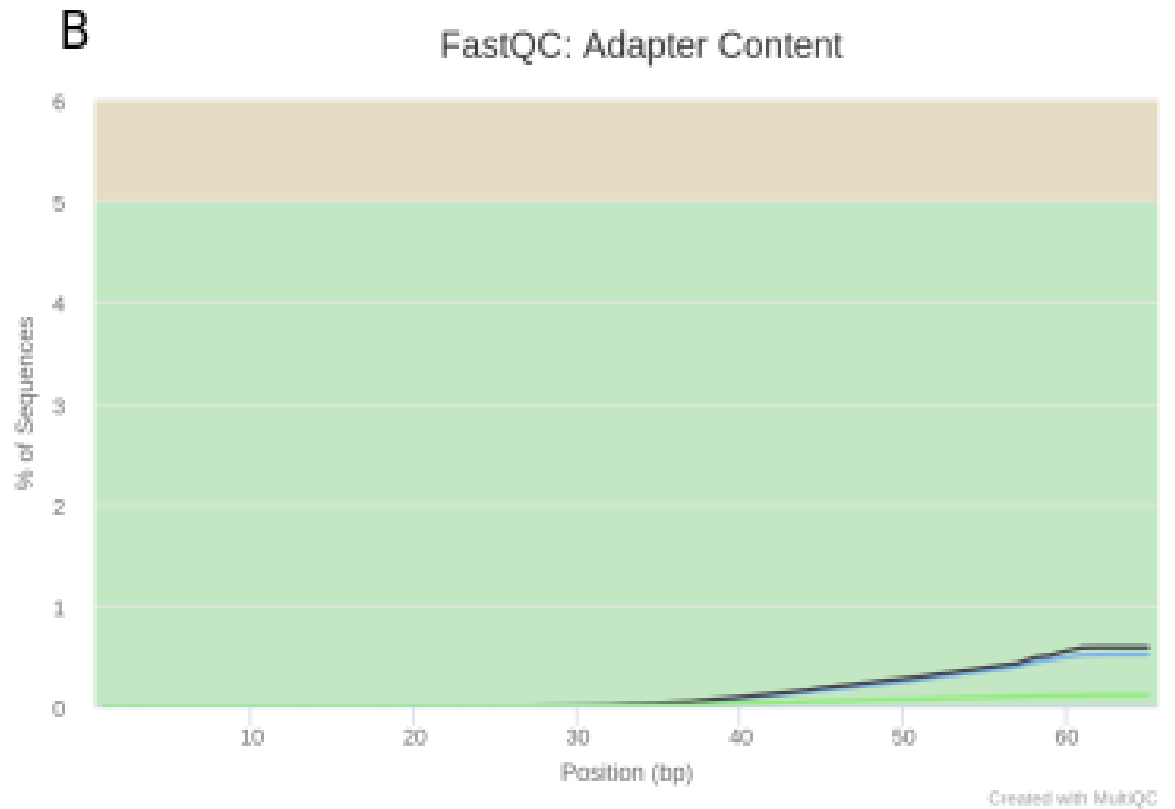


Figura Suplementar 1: Comparação de resultados da ferramenta FastQC. **A:** Qualidade média dos dados, onde as duas linhas laranjas representam os valores de qualidade dos datasets de Cq_albi. **B:** Conteúdo de adaptadores, onde as linhas azul e preta representam os datasets de Cq_albi. Fonte: O autor.

C. Tabelas suplementares

Dezordi, Filipe (2020), “Dissertação de Mestrado Acadêmico - Apêndice 3”, Mendeley Data, v1. Disponível em: <http://dx.doi.org/10.17632/kp96k2xxcm.1>.

D. Outras informações:

D.1 Disciplinas obrigatórias e eletivas:

Foi obtido aprovação em todas as disciplinas obrigatórias e eletivas com conceito A.

D.2 Participação em eventos:

III Simpósio Norte Nordeste de Bioinformática (ouvinte);

X-Meeting 2018 - 14th International Conference of the Brazilian Association of Bioinformatics and Computational Biology (apresentação em banner);

VII Semana de Biociências e Biotecnologia em Saúde - “Ciência na Praça: Que tal um café por 5 minutos de conversa?”, mesa redonda “Divulgação Científica” (palestrante);

VII Semana de Biociências e Biotecnologia em Saúde - “Evolução 101: Primeiros passos com análises filogenéticas”, minicurso (ministrante).

D.3 Produção bibliográfica:

DEZORDI, F. Z.; WALLAU, G. L.; Evolutionary history of a new group of Endogenous Viral Elements derived from Chuvirus in Mosquito Genomes: Just Junk-DNA or scars from an evolutive conflict? In: X-Meeting 2018 - 14th International Conference of the Brazilian Association of Bioinformatics and Computational Biology, 2018, São Pedro.

D.4 Cursos:

Curso de Boas Práticas na Pesquisa (OMS/IRR e Fapesp);

Theoretical and practical metagenomic approaches to viral Discovery (Universidade de São Paulo).

D.5 Scripts e Ferramentas de bioinformática:

Todos os *scripts* e linhas de comando para uso das ferramentas de bioinformática utilizados neste trabalho podem ser encontrados no link: <https://benchling.com/s/etr-5gCDVtteO7JRoKDzxO03>.