



Universidade Federal de Pernambuco
Centro de Biociências

GUTENBERGMANN BATISTA COUTINHO

**CARACTERIZAÇÃO FUNCIONAL *IN SILICO* DE
ELEMENTOS VIRAIS ENDOGENIZADOS DETECTADOS EM
GENOMAS DE CULICÍDEOS DE IMPORTÂNCIA MÉDICA**

Recife
2022

GUTENBERGMANN BATISTA COUTINHO

**CARACTERIZAÇÃO FUNCIONAL *IN SILICO* DE
ELEMENTOS VIRAIS ENDOGENIZADOS DETECTADOS EM
GENOMAS DE CULICÍDEOS DE IMPORTÂNCIA MÉDICA**

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação em
Biomedicina da Universidade Federal de
Pernambuco, como pré-requisito à
obtenção do título de Bacharel em
Biomedicina.

Orientador: Gabriel da Luz Wallau

Coorientador: Filipe Zimmer Dezordi

Recife
2022

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Coutinho, Gutembergmann Batista .

Caracterização funcional in silico de elementos virais endogenizados detectados em genomas de culicídeos de importância médica / Gutembergmann Batista Coutinho. - Recife, 2022.

64 : il., tab.

Orientador(a): Gabriel da Luz Wallau

Cooorientador(a): Filipe Zimmer Dezordi

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Biociências, Biomedicina, 2022.

Inclui referências, apêndices.

1. Elementos Virais Endogenizados. 2. Vírus. 3. Culicidae. 4. Elementos Transponíveis. I. Wallau, Gabriel da Luz. (Orientação). II. Dezordi, Filipe Zimmer. (Coorientação). III. Título.

570 CDD (22.ed.)

GUTENBERGMANN BATISTA COUTINHO

**CARACTERIZAÇÃO FUNCIONAL *IN SILICO* DE
ELEMENTOS VIRAIS ENDOGENIZADOS DETECTADOS EM
GENOMAS DE CULICÍDEOS DE IMPORTÂNCIA MÉDICA**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Biomedicina da Universidade Federal de Pernambuco, como pré-requisito à obtenção do título de Bacharel em Biomedicina.

Aprovada em: 06/06/2022

BANCA EXAMINADORA

Orientador: Dr Gabriel da Luz Wallau
Instituto Aggeu Magalhães/Fiocruz PE, Departamento de Entomologia

Msc. Alexandre Freitas da Silva
Instituto Aggeu Magalhães/Fiocruz PE, Departamento de Entomologia

Msc. João Luiz de Lemos Pitta
Instituto Aggeu Magalhães/Fiocruz PE/ Departamento de Microbiologia

Dedico este trabalho aos meus pais, que são meu maior exemplo de coragem, persistência e dedicação, valores essenciais para chegar até aqui.

AGRADECIMENTOS

À Deus, por tornar possível esse sonho, que por tanto tempo pareceu distante.

À minha família, especialmente aos meus pais, Mariluce e Onildo, por sempre estarem ao meu lado e não me permitir desistir.

À Universidade Federal de Pernambuco, por me mostrar que a educação é sempre o melhor caminho a ser seguido e me proporcionar tantas experiências essenciais na minha formação pessoal e profissional.

À equipe do Departamento de Entomologia, em especial ao Dr. Gabriel Wallau e ao Msc. Filipe Zimmer por toda a colaboração do início ao fim do projeto, essenciais para o seu desenvolvimento, e por buscar meios de realizá-lo, mesmo durante um período de pandemia, e a Yago Dias, por compartilhar a experiência de Iniciação Científica da melhor maneira possível.

À FACEPE por me proporcionar a experiência de vivenciar a Iniciação Científica, podendo assim retribuir à sociedade tudo que já me foi proporcionado.

Aos amigos de dentro e fora da universidade, que foram essenciais em todo esse processo e mostraram que até os momentos ruins nos trazem aprendizado. Gratidão a Celso, Ewerton, Igor, Karol, Manu, Rafael e Stephanie que acompanham minha caminhada desde os tempos remotos do colégio. À Angela, Akiria, Beatriz, Camila, Celina, Celine, Clara, Eduarda, Eliza, Emily, Guilherme, Gabriel, João Lucas, Laryssa, Lidiane, Lilian, Milena, Mirela e Wyrms por estarem presentes nessa infundável jornada acadêmica. À Arion, que muito além de um amigo foi um orientador em diversos momentos de minha graduação.

À professora Patrícia Dantas, que durante o Ensino Médio me apresentou a biologia da melhor maneira possível, o que me fez escolher a Biomedicina.

À equipe do PROEXPA por mostrar que podemos mudar a realidade das pessoas, mesmo que com pequenas atitudes e em especial à Dr^a. Janaína Rocha por me abrir as portas para participar desse belo projeto e auxiliar no sucesso de nossas ações.

Aos funcionários do Laboratório do Hospital das Clínicas e em especial aos estagiários Anna, Beatriz, Felipe, Gabriel, Guilherme, Marina, Rennaly,

Talita e Yan, por todo o conhecimento, competência e carinho, que fizeram meus dias serem mais proveitosos durante essa etapa tão crucial da graduação.

À ciência por me inspirar a resolver problemas e situações e me mostrar que nunca saberei de tudo mas sempre poderei dar o meu melhor.

À todos os funcionários do transporte público, que permitiram minha chegada e retorno das aulas, e da universidade, que auxiliaram no meu conforto e de outros milhares de estudantes diariamente, permitindo com que nossa rotina fosse mais agradável.

Por fim, também gostaria de agradecer a mim por ter seguido esse caminho, nunca esquecer do meu sonho e mostrar que sou capaz de realizá-lo.

"Eles não vão vencer, baby. Nada há de ser em vão."
(Johnny Hooker e Liniker)

COUTINHO, Gutembergmann Batista; WALLAU, Gabriel da Luz; DEZORDI, Filipe Zimmer. **Caracterização funcional *in silico* de elementos virais endogenizados detectados em genomas de culicídeos de importância médica.** 2022. 64 folhas. Trabalho de Conclusão de Curso (Graduação em Biomedicina) - Universidade Federal de Pernambuco, Recife, 2022.

RESUMO

Os elementos virais endogenizados (EVEs) são sequências de nucleotídeos originados a partir de infecções virais. Sua mobilização e replicação pode impactar no genoma do hospedeiro através de novas características benéficas ou danosas ao hospedeiro. Os mosquitos, como os do gênero *Aedes*, possuem uma grande quantidade de EVEs em seus genomas e esses têm sido associados principalmente a respostas antivirais. Outros gêneros, como o *Culex*, também possuem EVEs, porém em menor quantidade. Este trabalho tem como objetivo geral compreender a abrangência e dinâmica de EVEs em genomas de três espécies de mosquitos vetores: *Aedes aegypti*, *Aedes albopictus* e *Culex quinquefasciatus*. Para isso, foram avaliados 78 EVEs de cópia única e esses foram mapeados contra bibliotecas de sequenciamento de RNA (RNA-seq) para obter *reads*, de maneira quantitativa. Foram também analisadas as regiões flanqueadoras dos EVEs, para detecção de *Open Reading Frames* e as proteínas correspondentes a essas regiões. Por fim, foi feita uma filogenia utilizando as proteínas associadas às famílias virais dos respectivos EVE. A partir da contagem de *reads per kilobase per million* (RPKM) e avaliação dos mapeamentos, foram selecionados 15 EVEs, das famílias Flaviviridae, Chuviridae e Rhabdoviridae. Observou-se uma associação de EVEs com proteínas virais que resultou em diferentes padrões na profundidade do mapeamento. Nas filogenias, houve uma tendência de agrupamento em clados basais em alguns EVEs, indicando diferenças no processo evolutivo, comparado às proteínas virais. Os resultados sugerem ainda uma associação entre transposons na expressão de mRNA através do mecanismo de piRNAs.

Palavras-chave: Elementos Virais Endogenizados. Vírus. Culicidae. Elementos Transponíveis.

COUTINHO, Gutembergmann Batista; WALLAU, Gabriel da Luz; DEZORDI, Filipe Zimmer. **In silico functional characterization of endogenous viral elements detected in genomes of medically important Culicidae**. 2022. 64 pages. Completion of course work (Graduate in Biomedicine) – Federal University of Pernambuco, Recife, 2022.

ABSTRACT

Endogenized viral elements (EVEs) are nucleotide sequences originated from viral infections. Its mobilization and replication can impact the host genome through new beneficial or harmful characteristics to the host. Mosquitoes, such as those of the genus *Aedes*, have a large amount of EVEs in their genomes and these have been mainly associated with antiviral responses. Other genres, such as *Culex*, also have EVEs, but to a lesser extent. The general objective of this work is to understand the scope and dynamics of EVEs in the genomes of three species of vector mosquitoes: *Aedes aegypti*, *Ae. albopictus* and *Culex quinquefasciatus*. For this, 78 EVEs were evaluated and these were mapped against RNA sequencing libraries (RNA-seq) to obtain reads, in a quantitative way. The flanking regions of the EVEs were also analyzed for the detection of Open Reading Frames and the proteins corresponding to these regions. Finally, a phylogeny was made using the proteins associated with the viral families of the respective EVE. From the count of reads per kilobase per million (RPKM) and evaluation of the mappings, 15 EVEs were selected from Flaviviridae, Chuviridae and Rhabdoviridae families. An association of EVEs with viral proteins was observed that resulted in different patterns in the depth of mapping. In the phylogenies, there was a tendency of grouping in basal clades in some EVEs, indicating differences in the evolutionary process, compared to viral proteins. The results also suggest an association between transposons in mRNA expression through the piRNAs mechanism.

Key words: Endogenous Viral Elements. Virus. Culicidae. Transposable elements.

LISTA DE ILUSTRAÇÕES

Figura 1 – fêmea do mosquito adulto <i>Aedes Aegypti</i>	18
Figura 2 – fêmea do mosquito adulto <i>Aedes albopictus</i>	19
Figura 3 – fêmea do mosquito adulto <i>Culex quinquefasciatus</i>	20
Figura 4 – número amostral total e após a triagem, destacando a fase evolutiva das bibliotecas de RNA-seq e colônias selecionadas	32
Figura 5 – Distribuição do padrão de contagem de reads mapeados, por espécime, tecido, fase evolutiva e colônia, associado aos EVEs, para <i>Ae. aegypti</i>	35
Figura 6 - Distribuição do padrão de contagem de <i>reads</i> mapeados, por espécime, tecido, fase evolutiva e colônia, associado aos EVEs correspondentes, para <i>Ae. albopictus</i>	37
Figura 7 - Distribuição do padrão de contagem de <i>reads</i> mapeados, por espécime, tecido, fase evolutiva e colônia, associado ao EVE correspondente para <i>Culex. quinquefasciatus</i>	39
Figura 8 - EVE da família Flaviviridae, associado à proteínas virais, modificando a profundidade de mapeamento.	40
Figura 9 - EVE da família Chuviridae, associado à proteínas virais, modificando a profundidade de mapeamento.	40
Figura 10 - Filogenias associada aos EVEs da família viral Flaviviridae	42
Figura 11 - Filogenia de nucleoproteínas associada aos EVEs da família viral Rhabdoviridae	43
Figura 12.a - Filogenia de glicoproteína associadas aos EVEs da família viral Chuviridae	44
Figura 13.b - Filogenia de RdRP associada aos EVEs da família viral Chuviridae	45

LISTA DE TABELAS

Tabela 1 – Distribuição de EVEs, por família viral e tipo de amostra utilizada

33

LISTA DE ABREVIATURAS E SIGLAS

Ae.	<i>Aedes</i>
CFAV	Vírus agentes de fusão celular
Cx.	<i>Culex</i>
DNA	Ácido Desoxirribonucleico
dsRNA	RNA de fita dupla
enJSRVs	Betaretrovírus endógenos
EVE	Elemento viral endogenizado
ERV	Retrovírus endógenos
IGV	<i>Integrative Genomics Viewer</i>
iTOL	<i>Interactive Tree Of Life</i>
KRV	Vírus do Rio Kamiti
LTR	Longas repetições terminais/Long terminal repeats
LVP	Liverpool
MLV	Vírus da leucemia murina
MMTV	Vírus do tumor mamário de camundongo
mRNA	RNA mensageiro
NGS	Sequenciamento de Nova Geração
NCBI	<i>National Center for Biotechnology Information</i>
ORF	Quadro aberto de leitura/ <i>Open Reading Frame</i>
NIRV	Sequência de vírus de RNA integrado não retroviral
piRNA	RNAs que interagem com Piwi
RdRP	RNA polimerase dependente de RNA
RNA	Ácido ribonucleico
RNAi	RNAs de interferência
RNA-seq	Sequenciamento de RNA
RPKM	<i>Reads per kilobase per million</i>
SRA	<i>Sequence Read Archive</i>
sRNA	<i>Small RNA</i> /Pequenos RNA
ssRNA	RNA de fita simples
TE	Elementos transponíveis
vDNA	cDNA derivado de vírus

SUMÁRIO

1 INTRODUÇÃO	14
2 REVISÃO BIBLIOGRÁFICA	17
2.1 CULICÍDEOS	17
2.1.1 Família Culicidae	17
2.1.2 <i>Aedes aegypti</i> (Linnaeus, 1762)	17
2.1.3 <i>Aedes albopictus</i> (Skuse, 1894)	18
2.1.4 <i>Culex quinquefasciatus</i> (Say, 1823)	19
2.2 ELEMENTOS VIRAIS ENDOGENIZADOS	20
2.2.1 Histórico	20
2.2.2 Elementos Retrovirais e Não Retrovirais Endogenizados	22
2.2.3 EVEs em Genomas de Animais	23
2.2.4 EVEs em Mosquitos	24
2.3 BIOINFORMÁTICA	25
2.3.1 Montagem por referência (Bowtie2)	25
2.3.2 ORF Finder	26
2.3.3 Integrative Genomics Viewer (IGV)	26
2.3.4 KaryoploteR	27
3 OBJETIVOS	28
3.1 OBJETIVO GERAL	28
3.2 OBJETIVOS ESPECÍFICOS	28
4 METODOLOGIA	29
4.1 SELEÇÃO DE EVES	29
4.2 SELEÇÃO DAS BIBLIOTECAS DE RNA-SEQ	29
4.3. MAPEAMENTO E SELEÇÃO DE EVES PROMISSORES	29
4.4 ANÁLISE DOS FLANCOS	30
4.5 FILOGENIAS	31

	13
5 RESULTADOS	32
5.1 <i>AEDES AEGYPTI</i>	34
5.2 <i>AEDES ALBOPICTUS</i>	36
5.3 <i>CULEX QUINQUEFASCIATUS</i>	38
4.4 PROTEÍNAS	39
4.5 FILOGENIAS	41
6 DISCUSSÃO	46
7 CONCLUSÕES	50
REFERÊNCIAS	50
APÊNDICES	59
APÊNDICE A - DOMÍNIOS PROTEICOS PUTATIVOS CONSERVADOS	59
APÊNDICE B - MATERIAL SUPLEMENTAR	64

1 INTRODUÇÃO

Vírus são as entidades biológicas mais frequentes no mundo (KOONIN; DOLJA; KRUPOVIC, 2015; PAEZ-ESPINO *et al.*, 2016), estando presentes em todos os continentes. Possuem alto potencial adaptativo, podendo se estabilizar mesmo em condições extremas (CHAUHAN; GORDON, 2020; POLISCHUK *et al.*, 2007). Eles foram originalmente conceituados como “partícula causadora de doença que passou através de um filtro que retém todas as bactérias conhecidas” ou “*contagium vivum fluidum*”, pela capacidade de ultrapassar os menores filtros disponíveis na época (BEIJERINCK, 1898). Após mais de um século de estudos, a simplicidade desse conceito primordial é cada vez mais contrastante com tamanha complexidade viral.

Apesar de extremamente pequenos, geralmente variando entre 20 a 1000 nm (TORTORA; CASE; FUNKE, 2016, p. 360), os vírus possuem uma ampla diversidade de possíveis conformações de genomas, sendo os únicos “organismos” capazes de apresentar DNA de fita dupla ou simples, ou RNA, de fita simples ou dupla, de sentido positivo ou negativo (BERMAN, 2019). São frequentemente classificados como parasitas intracelulares obrigatórios, necessitando se associar a hospedeiros celulares para manter sua viabilidade e replicação. Apesar de abranger a maioria desses organismos, já foram relatados vírus que se mantêm viáveis fora das células hospedeiras (BERMAN, 2019; HÄRING *et al.*, 2005).

A relação do vírus e seu hospedeiro vai além da necessidade de se associar a um organismo. No caso dos retrovírus, vírus da família Retroviridae que possuem em sua etapa de replicação a transcrição reversa, e integração no genoma da célula hospedeira (COFFIN *et al.*, 2021), essa associação vai além: eles possuem a capacidade de integrar sequências genômicas no genoma de seus hospedeiros ao infectá-los, durante seu ciclo de replicação (WEISS, 2017). Tais regiões são denominadas elementos retrovirais endogenizados (ERVs, do inglês *Endogenous retroviruses*). Esse processo também pode ocorrer por meio de vírus não retrovirais de diversas famílias, utilizando mecanismos oferecidos por seus hospedeiros, formando os vírus de RNA não retroviral integrados (NIRVs, do inglês *non-retroviral integrated RNA virus sequences*) (FESCHOTTE; GILBERT, 2012; KATZOURAKIS; GIFFORD, 2010). NIRVs e ERVs são também conhecidos como Elementos Virais Endogenizados (EVEs).

Podem ocorrer em qualquer célula nucleada do hospedeiro, porém quando ocorrem em linhagens germinativas, são capazes de se integrar ao cromossomo, conseqüentemente levando à herança para seus descendentes por transmissão vertical, podendo, inclusive, ser fixada ao genoma da espécie hospedeira, dependendo da frequência em que ocorrem (FESCHOTTE; GILBERT, 2012; JOHNSON, 2019). O surgimento e manutenção dos EVEs pode ter um impacto neutro ou impactar positivamente e/ou negativamente os organismos, podendo ainda originar novos genes funcionais no hospedeiro (ARMEZZANI *et al.*, 2014; TER HORST *et al.*, 2019; ITO *et al.*, 2013).

EVEs estão presentes em diversos organismos. Em animais vertebrados, por exemplo, diversos processos podem ser resultantes desse mecanismo, como a associação de EVEs com genes embrionários, responsáveis pela morfogênese placentária e por impedir a rejeição do embrião através de mecanismos imunossupressores (BLAISE *et al.*, 2003b; DUPRESSOIR *et al.*, 2009). Também é relatada uma associação de EVEs com a patogênese do linfoma de Hodgkins (LAMPRECHT *et al.*, 2010).

Algumas espécies da família Culicidae, como *Aedes aegypti* e *Aedes albopictus* apresentam um número alto de NIRVs, que pode chegar a centenas (BLAIR; OLSON; BONIZZONI, 2020). Estudos também demonstram a presença de EVEs no *Culex quinquefasciatus*, como descrito por Russo *et al.* (2019), em que 28 NIRVs são relatados em genomas da espécie. Em culicídeos, uma das principais funções desenvolvidas por EVEs é a capacidade de influenciar o sistema imunológico e na competência vetorial. Esses elementos podem auxiliar nas respostas antivirais do próprio mosquito através da produção de RNAs de interferência (RNAi), como o Small RNA (sRNAs) e *PIWI-interacting RNA* (piRNAs) (PALATINI *et al.*, 2017; TASSETTO *et al.*, 2019; WHITFIELD *et al.*, 2017).

A produção dos RNAi leva à diminuição da replicação de vírus infectantes no organismo dos mosquitos, conseqüentemente causando modificações na susceptibilidade desses artrópodes à infecções virais (PALATINI *et al.*, 2017; TASSETTO *et al.*, 2019; WHITFIELD *et al.*, 2017). Através dessa via, utiliza-se uma sequência base do vírus para reconhecimento e formação de um complexo proteico que induz o silenciamento de sequências de RNA alvos do vírus, através da complementaridade entre a sequência viral e a dos EVEs (BRONKHORST; VAN RIJ, 2014; OLSON; BLAIR, 2015). A elucidação de novos mecanismos antivirais,

mediados por vírus endógenos, em mosquitos é de grande interesse científico, principalmente quando se trata de espécies de importância médica.

Além disso, os vírus, principalmente os de RNA, possuem a mais alta taxa evolutiva entre os organismos, podendo ser até um milhão de vezes mais rápida que a de seus hospedeiros celulares (AIEWSAKUN; KATZOURAKIS, 2015; HANADA; SUZUKI; GOJOBORI, 2004; KUMAR; SUBRAMANIAN, 2002; JENKINS *et al*, 2002; SANJUÁN, 2012), o que dificulta a compreensão aprofundada de seus perfis evolutivos, especialmente em longo prazo. Portanto, EVEs surgem como uma alternativa para guiar tais estudos, pois acompanham o ritmo evolutivo de seus hospedeiros celulares, permitindo estudos sobre evolução, origem e rotas de dispersão virais, descoberta de hospedeiros, até então desconhecidos ou de vírus extintos (AIEWSAKUN; KATZOURAKIS, 2015; DUFFY; SHACKELTON; HOLMES, 2008).

Compreender quais elementos são expressos à nível de RNA e os níveis em que a expressão ocorre, em diferentes EVEs e espécies, surge como uma proposta para guiar futuros estudos que busquem a avaliação do impacto funcional desses elementos na biologia dos hospedeiros, demonstrando quais os principais alvos que devem ser avaliados. Portanto, torna-se essencial a compreensão desses elementos e seu potencial de influenciar o genoma de seus hospedeiros, especialmente de mosquitos, por serem considerados os principais vetores de arboviroses para humanos (FANG, 2010).

2 REVISÃO BIBLIOGRÁFICA

2.1 CULICÍDEOS

2.1.1 Família Culicidae

A família Culicidae é um dos principais grupos de artrópodes, incluindo cerca de 3500 espécies e subespécies de mosquitos e seu nome é derivado do latim “*Culex*”, que significa “mosquito”. Podem ser subdivididas em duas subfamílias *Anophelinae* e *Culicinae* (FOSTER; WALKER, 2019). Estão geograficamente distribuídos por todos os continentes, exceto a Antártida, se adaptando principalmente a locais com água parada, essencial para o desenvolvimento das larvas (FOSTER; WALKER, 2019; ALMEIDA; COTA; RODRIGUES, 2020). As fêmeas possuem comportamento hematófago, pois necessitam do sangue para a maturação dos ovos (ROY *et al.*, 2016). Porém, durante o repasto sanguíneo, também atuam como vetores, transmitindo vírus causadores da dengue, febre amarela, chikungunya, Zika, filariose linfática e malária (FOSTER; WALKER, 2019; ROY *et al.*, 2016;)

2.1.2 *Aedes aegypti* (Linnaeus, 1762)

O *Aedes aegypti* (Figura 1), ou mosquito da febre amarela, é considerado o principal vetor da dengue, febre amarela, Chikungunya e Zika (FOSTER; WALKER, 2019; GUBLER, 2012; KOTSAKIOZI *et al.*, 2017). Esta espécie pode ser subdividida em duas subespécies: *formosus*, forma selvagem, de coloração preta, encontrado principalmente no interior da África, e *aegypti*, forma doméstica, de coloração marrom escura, distribuída principalmente nas Américas, sul da Ásia e regiões litorâneas da África (FOSTER; WALKER, 2019).

Originalmente descrito no Egito, foi introduzido no Brasil no período colonial, possivelmente devido ao tráfico de escravos, com fácil adaptação a regiões tropicais e subtropicais (CONSOLI; OLIVEIRA, 1994). Durante a década de 30, o mosquito era encontrado em uma área significativa no país, sendo responsável por uma intensa transmissão de febre amarela nas regiões Norte e Nordeste, o que motivou uma ação de combate ao mosquito e que levou à sua erradicação em 1955 (BRAGA; VALE, 2007; LÖWY, 1999). Porém, devido à sua presença em países

vizinhos e à descontinuidade das medidas de combate, esse foi sendo reintroduzido no país entre o final da década de 60 e durante a década de 70, sendo encontrado até os dias atuais e distribuído por todos os Estados (CONSOLI; OLIVEIRA, 1994; GUBLER, 2020; LIMA *et al.*, 2011).

Figura 1. fêmea do mosquito adulto *Aedes Aegypti*



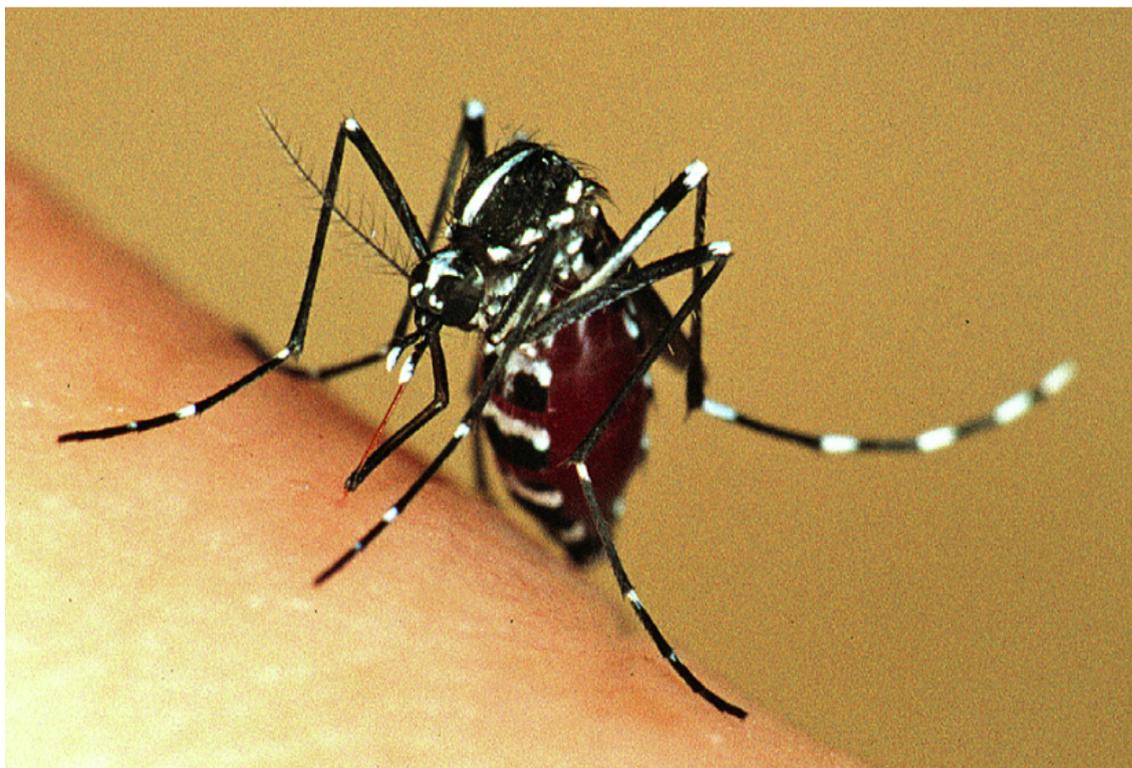
Fonte: FOSTER e WALKER (2019, p. 289)

2.1.3 *Aedes albopictus* (Skuse, 1894)

O *Ae. albopictus* (Figura 2), também conhecido como “tigre asiático”, é um vetor secundário da Dengue, febre amarela e Chikungunya (GUBLER, 2005; PAUPY *et al.*, 2018), podendo também ser associado ao Zika vírus (GARDNER, 2016). Esse mosquito pode coexistir em ambientes urbanos, periurbanos, rurais e silvestres (DOS SANTOS *et al.*, 2018). É proveniente do sudeste asiático, originalmente descrito na Índia, e foi relatado pela primeira vez no Brasil, no estado do Rio de Janeiro, em 1986, possivelmente introduzido através de portos localizados no Espírito Santo, sendo disseminado por meio de ferrovias do Vale do Rio Doce (CONSOLI; OLIVEIRA, 1994; FORATTINI, 1986). Atualmente, essa espécie se disseminou pelo país e está presente em cerca de 60% dos municípios

brasileiros, sendo uma espécie que requer atenção para possíveis futuras epidemias (DOS SANTOS *et al.*, 2018; PANCETTI, 2015).

Figura 2: fêmea do mosquito adulto *Aedes albopictus*



Fonte: FOSTER e WALKER (2019, p. 292)

2.1.4 *Culex quinquefasciatus* (Say, 1823)

O *Cx. quinquefasciatus* (Figura 3), também conhecido como mosquito doméstico tropical do sul, pernilongo ou muriçoca, é um importante vetor de patógenos, incluindo o *Wuchereria bancrofti* (BRASIL, 2011; MORAIS; MARRELLI; NATAL, 2006), causador da filariose linfática, vírus emergentes como o vírus do Nilo Ocidental (LIMA-CAMARA, 2016). Estudos experimentais também apontam que o Zika virus pode ser infectado e atuar como transmissor do Zika vírus (GUEDES *et al.*, 2017; AYRES *et al.*, 2019). Foi inicialmente relatado nos Estados Unidos mas também é relatado nas África, Ásia, Américas e Oceania (FOSTER; WALKER, 2019). Entre os anos de 1951 e 1955, ocorreram as primeiras ações específicas de combate ao *Cx. quinquefasciatus* no Brasil, para controlar a filariose linfática (MEDEIROS *et al.*, 2003).

Este mosquito é altamente antropofílico, sendo encontrado em área urbana ou rural, e raramente relatado em locais com ausência de humanos. Está associado principalmente a hábitos noturnos, horário em que ocorre o estímulo quanto à hematofagia (CONSOLI; OLIVEIRA, 1994). Apesar do Brasil ter interrompido a transmissão da filariose linfática no ano de 2013, a doença ainda não foi oficialmente erradicada, fazendo-se necessária a vigilância do mosquito (BRASIL, 2021).

Figura 3: fêmea do mosquito adulto *Culex quinquefasciatus*



Fonte: VIEIRA; FIOCRUZ, 2016

2.2 ELEMENTOS VIRAIS ENDOGENIZADOS

2.2.1 Histórico

O primeiro relato de endogenização viral documentado é datado à década de 60, quando ocorria um surto de leucose em frangos pelo vírus da leucose aviária endógena (ALV). Houve a necessidade da realização de um teste sorológico de fixação do complemento, denominado COFAL, para detecção de proteínas gag virais (WEISS, 2006). Houve estranhamento quanto ao resultado pelo fato de que

muitas galinhas, aparentemente não infectadas, apresentaram positividade, levando à uma insegurança quanto à sensibilidade do teste (DOUGHERTY; STEFANO, 1966). Em seguida, sugeriu-se que a herança gênica do antígeno associado ao Gag como um gene mendeliano dominante após ocorrerem cruzamentos entre galinhas Gag-positivas e negativas (PAYNE; CHUBB, 1968).

Apesar de na época a hipótese de retrovírus transferirem conteúdo gênico para os hospedeiros ser inviável, posteriormente a descoberta da enzima transcriptase reversa como capaz de converter o RNA em DNA levou à hipótese como uma real possibilidade (BALTIMORE, 1970; MAGER; STOYE, 2021; TEMIN; MIZUTANI, 1970). Na mesma época, estudos já demonstravam que a hibridização líquida podia identificar sequências associadas a retrovírus em DNA genômico (BENVENISTE; TODARO, 1974) e que a radiação e substâncias químicas poderiam induzir retrovírus em células que aparentemente não tinham nenhuma carga viral (LOWY *et al.*, 1971). Foram, então, detectados retrovírus endógenos (ERVs) associados aos vírus possivelmente cancerígenos, o vírus do tumor mamário de camundongo (MMTV) e vírus da leucemia murina (MLV), levando a hipótese da associação entre ERVs e o câncer humano (MAGER; STOYE, 2021);

Em seguida, foram utilizadas metodologias com abordagens biológicas, com o intuito de detectar a replicação viral, e estruturais, para encontrar sequências de ácido nucléico associadas a retrovírus. Um desses testes propôs tratar células com agentes ativadores de ERVs e cultura celular ou do sobrenadante livre dessas células com um indicador de marcadores virais, detectando, por exemplo, a atividade da transcriptase reversa (KHAN *et al.*, 2009). Dessa maneira, foram encontrados ERVs infecciosos em diversas espécies, mas não em humanos (KHAN *et al.*, 2009; MAGER; STOYE, 2021).

Como exemplo de abordagens estruturais, realizaram-se experimentos para clonagem de ERVs humanos com hibridização de baixo rigor, utilizando ERVs de outras espécies, sendo clonados vírus murinos e de galinha (CALLAHAM *et al.*, 1982; MARTIN *et al.*, 1981). Posteriormente, a técnica de PCR foi empregada, utilizando primers complementares a determinadas regiões selecionadas. Os resultados desses testes evidenciaram a diversidade de ERVs no DNA de diferentes espécies (DUNWIDDIW *et al.*, 1986; MEDSTRAND; BLOMBERG, 1993; SHIH; MISRA; RUSH, 1989; TRISTEM *et al.*, 1996;). Com o sequenciamento do genoma humano, foi possível observar que 8% do DNA é composto por ERVs

(CONSORTIUM, 2001; CONSORTIUM; 2002; CONSORTIUM, 2004; CONSORTIUM, 2005).

Em 2004, houve o primeiro relato de integração de sequências não retrovirais em genomas eucariotos. A inesperada descoberta trata-se da integração de quatro regiões semelhantes aos genomas de RNAs de vírus agentes de fusão celular (CFAV, do Inglês, *Cell fusion agent virus*) e do vírus do Rio Kamiti (KRV), sequências de RNA associadas à vírus do gênero flavivirus, em genomas de *Ae. albopictus* e *Ae. aegypti* (CROCHU, 2004).

O desenvolvimento de métodos analíticos *in silico* também ampliou as possibilidades de realização de novos estudos (LERAT, 2009), que somados aos métodos já existentes, permitem análises complexas de diferentes espécies, ampliando cada vez mais as possibilidades de identificação dessas sequências.

2.2.2 Elementos Retrovirais e Não Retrovirais Endogenizados

Os EVEs são derivados principalmente de retrovírus uma vez que a integração no genoma do hospedeiro é obrigatoriamente parte de seu ciclo de replicação (KATZOURAKIS; GIFFFORD, 2010). Porém, outras famílias virais também podem ser integradas através de etapas incomuns no ciclo de vida viral, como apresentado em Holmes (2011). O autor apresenta dois possíveis mecanismos que podem levar a endogenização de um vírus de RNA não retrovírus: O primeiro ocorre quando parte do genoma viral chega ao núcleo de células do hospedeiro, quando o normal é que este se limite a atingir o citoplasma. Esse mecanismo ainda é pouco compreendido; O outro meio é através da conversão de ssRNA em dsRNA. Holmes (2011) também demonstra que essa conversão pode ocorrer através de algumas etapas: a mais provável envolve a utilização da transcriptase reversa presente em retroelementos celulares frequentemente encontrados em eucariotos. Outra possibilidade compreende na recombinação não homóloga entre um retrotransposon e um vírus de RNA exógeno. Essa etapa pode acarretar na transcrição reversa e na integração do RNA viral (GEUKING *et al.*, 2009). O autor ainda cita a possibilidade de uma infecção simultânea entre diferentes vírus, sendo um deles um retrovírus que fornecerá a transcriptase reversa, resultando num ambiente com um alto número da enzima, que pode se associar a outro vírus presente.

2.2.3 EVEs em Genomas de Animais

Uma importante associação de EVEs com processos evolutivos em animais vem sendo demonstrada. Através da inserção de regiões genômicas virais, pode ocorrer o surgimento de genes exaptados, ou seja, genes responsáveis por desempenhar novas funções no organismo (KATZOURAKIS; GIFFFORD, 2010). Apesar de em muitos casos essas inserções serem perdidas em algumas gerações, há exemplos notórios de como essa complexa relação pode influenciar completamente a evolução de espécies diversas.

Os genes sincitina-1 e sincitina-2 estão presentes em genomas de humanos e primatas (BLAISE *et al.*, 2003). Já seus homólogos, mas não ortólogos, sincitina-A e sincitina-B encontram-se em linhagens de camundongos (DUPRESSOIR *et al.*, 2005; DUPRESSOIR *et al.*, 2009). Ambos os genes possuem significativa similaridade com envelopes virais (BLAISE *et al.*, 2003a; BLAISE *et al.*, 2003b; DUNLAP *et al.*, 2006; MI *et al.*, 2000). A expressão desses genes ocorre especificamente no sinciciotrofoblasto. Há evidências da atividade fusogênica, essencial na morfogênese da placenta nestes animais e de uma possível atuação como imunossupressores, essencial para impedir a rejeição do embrião (BLAISE *et al.*, 2003b; DUPRESSOIR *et al.*, 2009).

A possível associação entre EVEs e o desenvolvimento embrionário também é relatada através dos *endogenous betaretroviruses* (enJSRVs), um ERV expresso no trato genital e nas células mononucleares do trofotoderma do feto, embrião e de membranas extra embrionárias associadas à ovelhas (DUNLAP; PALMARINI; SPENCER, 2006a; DUNLAP *et al.*, 2006b). O envelope enJSRVs é encontrado em associação ao embrião durante a gestação desses animais e demonstra uma importante contribuição no desenvolvimento placentário, podendo evolutivamente ter melhorado a eficiência reprodutiva e o desenvolvimento embrionário e fetal (DUNLAP; PALMARINI; SPENCER, 2006a; DUNLAP *et al.*, 2006b).

Os efeitos benéficos também podem ser associados à proteção contra doenças, como é o caso do produto do gene de restrição do retrovírus murino, Fv1 (YAP *et al.*, 2014). O Fv1 é derivado de domínios gag provenientes de ERVs ancestrais, sendo alguns destes associados ao vírus da leucemia murina (MLV). Nestes animais, é relatado um efeito protetor contra a infecção do MLV, além de outros retrovírus, como de lentivírus e spumavirus (YAP *et al.*, 2014).

Um grande número de EVEs vem sendo encontrado em genomas de animais. Hayard *et al.* (2013) detectaram 87.750 retrovírus endógenos ao analisar 60 genomas de vertebrados. Com o desenvolvimento de novas metodologias, esse número tende a crescer, expandindo cada vez mais o conhecimento quanto à presença e impacto dos fragmentos virais nos genomas dos organismos.

2.2.4 EVEs em Mosquitos

Para um mosquito ser classificado como “vetor” de arboviroses, é preciso que ele tenha “competência vetorial” para tal. Ou seja, é necessário que ele possa ser infectado e que tenha capacidade de infectar outros organismo, suportando os seguintes estágios: tolerar a infecção viral pelas células epiteliais presentes no intestino, permitir a replicação do vírus, a disseminação deste para a hemocele e que as glândulas salivares possam ser infectadas, para o vírus ser, então, transmitido para outros organismos (BEERNTSEN, B. T. *et al.*; SALAZAR *et al.*, 2007; SEVERSON; BEHURA, 2016). Os EVEs podem influenciar na competência vetorial e no sistema imunológico de mosquitos através da produção dos *RNAs interference* (RNAi), principalmente *PIWI-interacting RNA* (piRNA), além de *Small RNA* (sRNA) (PALATINI *et al.*, 2017; TASSETO *et al.*, 2018; WHITFIELD *et al.* 2017).

Ae. aegypti e *Ae. albopictus* possuem um alto número de *Non-retroviral integrated RNA virus sequences* (NIRVs), os EVEs provenientes de famílias não retrovirais, em seus genomas, podendo chegar a centenas (BLAIR; OLSON; BONIZZONI, 2020). Já para o *Cx. quinquefasciatus*, Russo *et al.* (2019) relatam a presença de 28 NIRVs em genomas da espécie. A infecção viral em mosquitos pode promover a conversão de regiões de RNA viral em *viral derived* cDNA (vDNA) com a ação de transcriptases reversas (TASSETO, *et al.*, 2019). Goic *et al.*, (2016) demonstram que o vDNA também pode estar associado à competência vetorial de mosquitos.

No *Ae. aegypti* e *Ae. albopictus*, existe uma maior concentração de NIRVs em loci geradores de piRNA, os chamados clusters de piRNA, que podem controlar os elementos transponíveis presentes em células germinativas de seus hospedeiros (TER HORST *et al.*, 2019; PALATINI *et al.*, 2017; WHITFIELD *et al.*, 2017). A integração de NIRVs estrategicamente nessas regiões pode sugerir uma

associação de EVEs com mecanismos antivirais (OLSON; BONIZZONI, 2017). Possivelmente, os piRNAs podem utilizar EVEs como moldes durante sua formação, podendo, esses, até certo ponto oferecer memória imunológica para seu hospedeiro (MIESEN; JOOSTEN; VAN RIJ, 2016), alterando assim a competência vetorial para futuras infecções (PALATINI *et al.*, 2019; TASSETO *et al.*, 2019; WHITFIELD *et al.*, 2017).

Os NIRVs também podem estar presentes em regiões enriquecidas com elementos transponíveis (TE). Palatini *et al.* (2017) demonstram uma maior ocupação de TEs, principalmente de retrotransposons com *long terminal repeats* (LTR) e dos LTR da família Ty3_gypsy, em regiões com NIRVs, em comparação ao genoma completo de *Ae. aegypti* e *Ae. albopictus*. Os autores também demonstram a atividade transcricional de EVEs, que podem incorporar *Open reading frames* ao serem integrados em seus hospedeiros.

2.3 BIOINFORMÁTICA

2.3.1 Montagem por referência (Bowtie2)

O Bowtie2 é um *software* utilizado para realizar o mapeamento de sequências com genomas de referências. Diante da necessidade de se realizarem alinhamentos longos de maneira mais acessível, o Bowtie2 torna-se uma opção viável, visto que outras ferramentas existentes necessitam de muitos recursos computacionais e tempo para executar a função, podendo ainda gerar custos altíssimos para o estudo (LANGMEAD *et al.*, 2009).

A ferramenta permite a realização de alinhamentos de milhões de *reads*, mais rapidamente que a primeira geração do Bowtie (LANGMEAD *et al.*, 2009). O programa permite o alinhamento de sequências longas de até 1000 caracteres (BOWTIE, 2018?). Outra importante propriedade que o Bowtie2 dispõe é a capacidade de realizar alinhamentos com lacunas, que podem representar inserções e exclusões ou ser resultantes de erros, limitação imposta a algumas outros *softwares*, incluindo o Bowtie (LANGMEAD *et al.*, 2009). Outra importante característica é a alta taxa de sensibilidade e precisão do programa, mesmo em sequências incorretas (LANGMEAD; SALZBERG, 2012; LINDNET; FRIEDEL, 2012).

2.3.2 ORF Finder

Uma *Open Reading Frame* (ORF) é um “potencial gene codificador de proteínas” e pode ser definida como “uma sequência que tem um comprimento divisível por três e é limitada por códons de parada” (CLAVERIE, 1997; FERMIN *et al.*, 2006; SIEBER; PLATZER; SCHUSTER, 2018; WOODCROFT *et al.*, 2016). O *Open Reading Frame Finder* (ORF Finder), por sua vez, é um *software* oferecido pelo *National Center for Biotechnology Information* (NCBI) utilizado para a pesquisa de ORFs, disponibilizadas na plataforma do NCBI, em sequências enviadas pelo usuário (PATEL; PANCHAL; ANJARIA, 2012). Na ferramenta, tais ORFs podem ser filtradas de acordo com o comprimento mínimo, códon de início da ORF, além da possibilidade de definir o código genético utilizado na pesquisa. As proteínas retornadas na pesquisa podem ainda ser pesquisadas através das ferramentas SMART Blast e BLASTp (NCBI, 2004).

2.3.3 Integrative Genomics Viewer (IGV)

O *Integrative Genomics Viewer* (IGV) é um programa computacional desenvolvido em Java, de alta performance, que permite visualizar e explorar diferentes dados genômicos como leituras alinhamentos de sequências, expressão gênica, mutações, metilação, número de cópias, visualização de RNAi e anotações genômicas (ROBISON *et al.*, 2011; THORVALDSDÓTTIR; ROBINSON; MERISOV, 2012). O IGV ganha importância diante da grande quantidade de dados fornecidos com a utilização do Sequenciamento de Nova Geração (NGS), que pode chegar a gigabytes e até terabytes, podendo se tornar uma limitação em determinados estudos, mas que são suportados no *software* e, portanto, podem ser explorados de forma leve, rápida e intuitiva (ROBISON *et al.*, 2011; THORVALDSDÓTTIR; ROBINSON; MERISOV, 2012). A ferramenta permite a visualização de até milhares de dados simultaneamente, em diferentes escalas, desde pares de bases até o genoma completo (ROBISON *et al.*, 2011; THORVALDSDÓTTIR; ROBINSON; MERISOV, 2012).

2.3.4 KaryoploteR

O KaryoploteR é um pacote da linguagem R/Bioconductor utilizado para o desenvolvimento de *karyoplots*, ou seja, dados genômicos dispostos no genoma que podem ser visualizados linearmente e em conjunto. Dessa maneira, é possível observar os padrões dispostos ao longo das sequências e explorar dados complexos, sendo uma importante etapa em análises genômicas (GEL; SERRA, 2017). Alguns programas para visualização de dados possuem limitações, como o Circos, que apesar da alta capacidade de personalização, gera apenas ideogramas de layout circular, (KRZYWINSKI *et al.*, 2009), ou ggbio, que não dispõe de muitas opções de personalização (YIN; COOK; LAWRENCE, 2012).

Portanto, o karyoploteR surge para atender a demanda da criação de ideogramas não circulares altamente personalizáveis em qualquer organismo presente no *database*, inclusive podendo criar genomas personalizados para organismos indisponíveis. Esta biblioteca permite a plotagem de dados de baixo nível, para representar elementos gráficos simples, de maneira flexível, ou de dados de alto nível, que podem seguir conceitos biológicos, para a apresentação de dados mais complexos (GEL; SERRA, 2017).

3 OBJETIVOS

3.1 OBJETIVO GERAL

Compreender a abrangência e dinâmica de expressão de elementos virais endogenizados em bibliotecas de RNA-seq de *Aedes aegypti*, *Ae. albopictus* e *Culex quinquefasciatus*.

3.2 OBJETIVOS ESPECÍFICOS

- Avaliar o potencial de expressão de RNA mensageiro (mRNA) dos EVEs por metodologia *in silico*;
- Associar diferentes padrões de *reads* mapeados e associá-los a transposons e domínios proteicos;
- Realizar uma análise filogenética de proteínas transcritas pelo EVEs e proteínas da mesma família viral.
- Selecionar EVEs para futuras análises *in vitro* e *in vivo*.

4 METODOLOGIA

4.1 SELEÇÃO DE EVES

Apenas EVES de cópia única, disponibilizados em um *database* do grupo de pesquisa (DEZORDI, 2020), foram selecionados. Foram utilizados genomas de 3 diferentes espécies: *Ae. aegypti*, *Ae. albopictus* e *Cx. quinquefasciatus*, resultantes de diferentes montagens.

4.2 SELEÇÃO DAS BIBLIOTECAS DE RNA-SEQ

Os dados de sequenciamento de RNA (RNA-seq), das três espécies selecionadas, foram obtidos através do *Sequence Read Archive* (SRA - NCBI) (LEINONEN; SUGAWARA, 2010), sem haver distinção quanto aos tecidos selecionados. Apenas experimentos em condições classificadas como “controle”, ou seja, sem haver indução a situações de estresse, amostras *in vivo* com fase de desenvolvimento definida ou estudos realizados por metodologia *in vitro* foram selecionados.

4.3. MAPEAMENTO E SELEÇÃO DE EVES PROMISSORES

Os dados de RNA-seq recuperados no SRA na pesquisa foram cruzados com os respectivos EVES através da ferramenta Bowtie2 (LANGMEAD; SALZBERG, 2012), sendo os parâmetros otimizados para a realização de um mapeamento não-restritivo, considerando possíveis diferenças entre as linhagens dos dados de RNA-seq e montagens. Utilizaram-se os seguintes parâmetros:

```
--local --no-unal -D 20 -R 3 -L 3 -N 1 --gbar 1 --mp 3'
```

--local: alinhamento local

--no-unal: remoção de registros SAM para *reads* não alinhados

-D <int>: Podem haver falhas em até 20 tentativas consecutivas de extensão de *seeds*, ou sementes (nome dado a regiões idênticas ou quase idênticas entre

diferentes sequências) antes que o Bowtie 2 dê continuidade, usando alinhamentos encontrados até aquele momento.

-R <int>: Até 3 vezes no máximo que o Bowtie 2 poderá realizar uma "resemeadura" em *reads* com *seeds* repetitivas.

-L <int>: Define o comprimento das subsequências de *seeds* durante o alinhamento de várias delas. Com a diminuição do valor definido, o alinhamento se torna mais lento e sensível.

-N <int>: Estipula o número de incompatibilidades permitidas em um alinhamento de *seeds* quando várias delas estão sendo alinhadas.

--gbar <int>: Determina as posições em que não serão permitida a presença de lacunas, do início ou fim do *read*.

--mp MX,MN: Estabelece as penalidades máxima (MX) e mínima (MN) para *mismatches*.

Foi realizada uma contagem dos *reads* mapeados através do programa *Samtools* (Li *et al.*, 2009). A análise foi complementada com a ferramenta *Integrative Genomics Viewer* (IGV) (ROBINSON *et al.*, 2011), para visualização do mapeamento das sequências. Também utilizou-se o pacote *karyoploteR* (GEL; SERRA, 2017) para observar a expressão gênica destas. Após obter os resultados, os EVEs com maior contagem de *reads per kilobase per million* (RPKM), calculado através da estimativa da profundidade de mapeamento das bases dividido pelo tamanho total da região, foram selecionados para estudos posteriores *in vivo*. Esses também passaram por uma triagem através da visualização de padrões utilizando o IGV. A cobertura do mapeamento foi utilizada como critério seletivo e apenas EVEs mapeados em toda, ou na maior parte de sua extensão, com homogeneidade na profundidade ao longo do EVE foram selecionados, sendo excluídos os que apresentavam cobertura insatisfatória ou excesso de possíveis eventos mutacionais.

4.4 ANÁLISE DOS FLANCOS

Foi realizada uma análise das *Open reading frames* com a ferramenta ORFfinder - NCBI, na versão web. Foram consideradas as proteínas localizadas tanto nas regiões flanqueadoras quanto nos EVEs, em até 10 kbp do EVE. Para serem selecionadas, as ORFs deveriam apresentar comprimento superior à 700 nt

e/ou estarem localizadas de modo adjacente, em até 1 kbp da região do EVE. O valor de RPKM do EVE foi dividido pelo RPKM das regiões flangeadoras *down* e *up* e pela soma de ambas, para assim classificar o padrão de expressão entre elas.

4.5 FILOGENIAS

Através da ferramenta ORF finder - NCBI (WHEELER *et al.*, 2001), foram pesquisadas as ORFs presentes nas sequências de EVEs selecionadas após as análises. Como filtro, apenas ORFs com mais de 300 nucleotídeos de comprimento, presentes na região do EVE, foram selecionadas, sendo excluídas aquelas de menor comprimento ou dispostas apenas nos flancos. As proteínas associadas às respectivas ORFs foram pesquisadas através do programa blastp, sendo aplicado o filtro para permitir a busca apenas de proteínas associadas às famílias virais dos respectivos EVE, presentes no banco de dados da ferramenta.

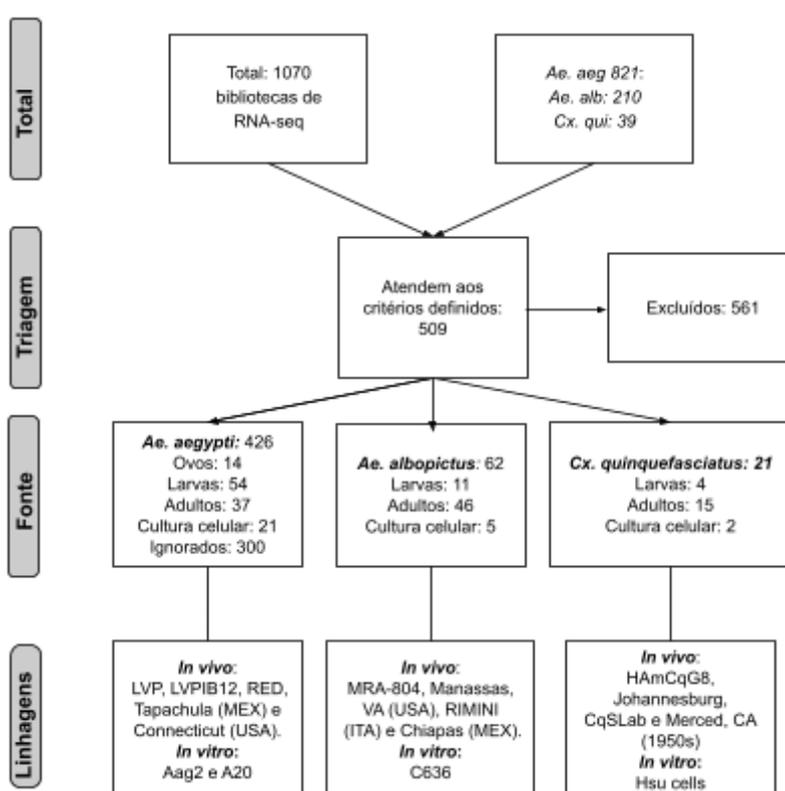
Entre as proteínas retornadas no blastp, apenas aquelas que apresentavam identidade maior ou igual à 25% foram selecionadas. Foram realizados alinhamentos múltiplos locais das proteínas com os EVEs na ferramenta Mafft *alignment* (KATOH; ROZEWICKI; YAMADA, 2019; KURAKU *et al.*, 2013), seguindo os parâmetros pré definidos no *software*. As sequências foram plotadas em arquivos, divididos por família viral e tipo de proteína. Os resultados foram editados através do programa AliView (LARSSON, 2014). A partir da versão final dos alinhamentos, as árvores filogenéticas foram reconstruídas, utilizando a ferramenta IQ-TREE, na versão web (MINH *et al.*, 2020), através da máxima verossimilhança. Foram utilizados os métodos bootstrap *ultrafast* e aLRT, com 1000 e 100 réplicas, respectivamente. Por fim, as filogenias foram plotadas na *Interactive Tree Of Life* (iTOL) (LETUNIC; BORK 2007; LETUNIC; BORK 2019) para visualização, anotações e edições.

5 RESULTADOS

Ao total, 1070 bibliotecas de RNA-seq das três espécies estudadas foram retornadas na pesquisa, sendo 821 de *Ae. aegypti*, 210 de *Ae. albopictus* e 39 de *Cx. quinquefasciatus* (Figura 4). Destas, 509 atendiam aos critérios de seleção do estudo, sendo estes, experimentos *in vitro* ou *in vivo* com o estágio de desenvolvimento definido e em condições de controle.

Para a espécie *Ae. aegypti*, foram selecionadas 426 bibliotecas de RNA-seq, com 37 dessas em fase adulta, 54 larvas, 14 ovos e 21 de linhagens celulares. Por apresentar um número amostral muito superior ao das demais espécies, 300 amostras de *Ae. aegypti* não foram avaliadas, sendo essa exclusão feita aleatoriamente, pois o número amostral obtido já era suficiente para os objetivos do estudo. Também foram selecionadas 62 bibliotecas da espécie *Ae. albopictus*, sendo 46 em fase adulta, 11 larvas e 5 de cultura celular. Por fim, para *Cx. quinquefasciatus*, foram selecionadas 21 bibliotecas, ao total, sendo 15 de mosquitos adultos, 4 larvas e 2 de linhagem celular.

Figura 4: número amostral total e após a triagem, destacando a fase evolutiva das bibliotecas de RNA-seq e colônias selecionadas



Dos 518 EVEs presentes no banco de dados do grupo, 259 são da espécie *Ae. aegypti*, 250 da *Ae. albopictus* e 9 da *Cx. quinquefasciatus*. Para *Ae. aegypti*, 38 EVEs eram de cópia única, atendendo, portanto, ao critério de seleção estabelecido. Desses, 20 pertencem à família Rhabdoviridae, 12 à Flaviviridae, 5 à Xinmoviridae e 1 à Chuviridae. Quanto à *Ae. albopictus*, 39 eram de cópia única, sendo 19 da família Flaviviridae, 13 da Rhabdoviridae e 7 da Chuviridae. Já para o *Cx. quinquefasciatus*, apenas 1 EVE da família Chuviridae foi selecionado para o estudo.

Após as análises, os EVEs que apresentaram resultados mais satisfatórios, de acordo com os critérios de análise estabelecidos, foram selecionados à parte, para futuros estudos funcionais *in vitro*. Para *Ae. aegypti*, 4 EVEs da família Rhabdoviridae, 2 da Flaviviridae e 1 da Chuviridae foram selecionados. Eles apresentaram um mapeamento de maior profundidade quando expressos em bibliotecas de adultos, além de cultura celular, também observado em Rhabdoviridae. Nenhum EVE da família Xinmoviridae apresentou cobertura satisfatória (apêndice B). Para *Ae. albopictus*, foram selecionados 4 EVEs da família Flaviviridae, 2 da Rhabdoviridae e 1 da Chuviridae, ambos expressos em cultura celular. Para *Cx. quinquefasciatus*, o único EVE, da família Chuviridae, mostrou resultados promissores em algumas das bibliotecas de fase adulta (Tabela 1).

Tabela 1. Distribuição de EVEs, por família viral e tipo de amostra utilizada

Espécie	Família viral (EVE)	EVEs (total)	EVEs selecionados	Expressão
<i>Ae. aegypti</i>	Rhabdoviridae	20	4	Adultos Cultura celular
<i>Ae. aegypti</i>	Flaviviridae	12	2	Adultos
<i>Ae. aegypti</i>	Chuviridae	1	1	Adultos
<i>Ae. aegypti</i>	Xinmoviridae	5	-	-
<i>Ae. albopictus</i>	Flaviviridae	19	4	Cultura celular
<i>Ae. albopictus</i>	Rhabdoviridae	13	2	Cultura celular
<i>Ae. albopictus</i>	Chuviridae	7	1	Cultura celular
<i>Cx. quinquefasciatus</i>	Chuviridae	1	1	Adulto

Fonte: elaborada pelos autores

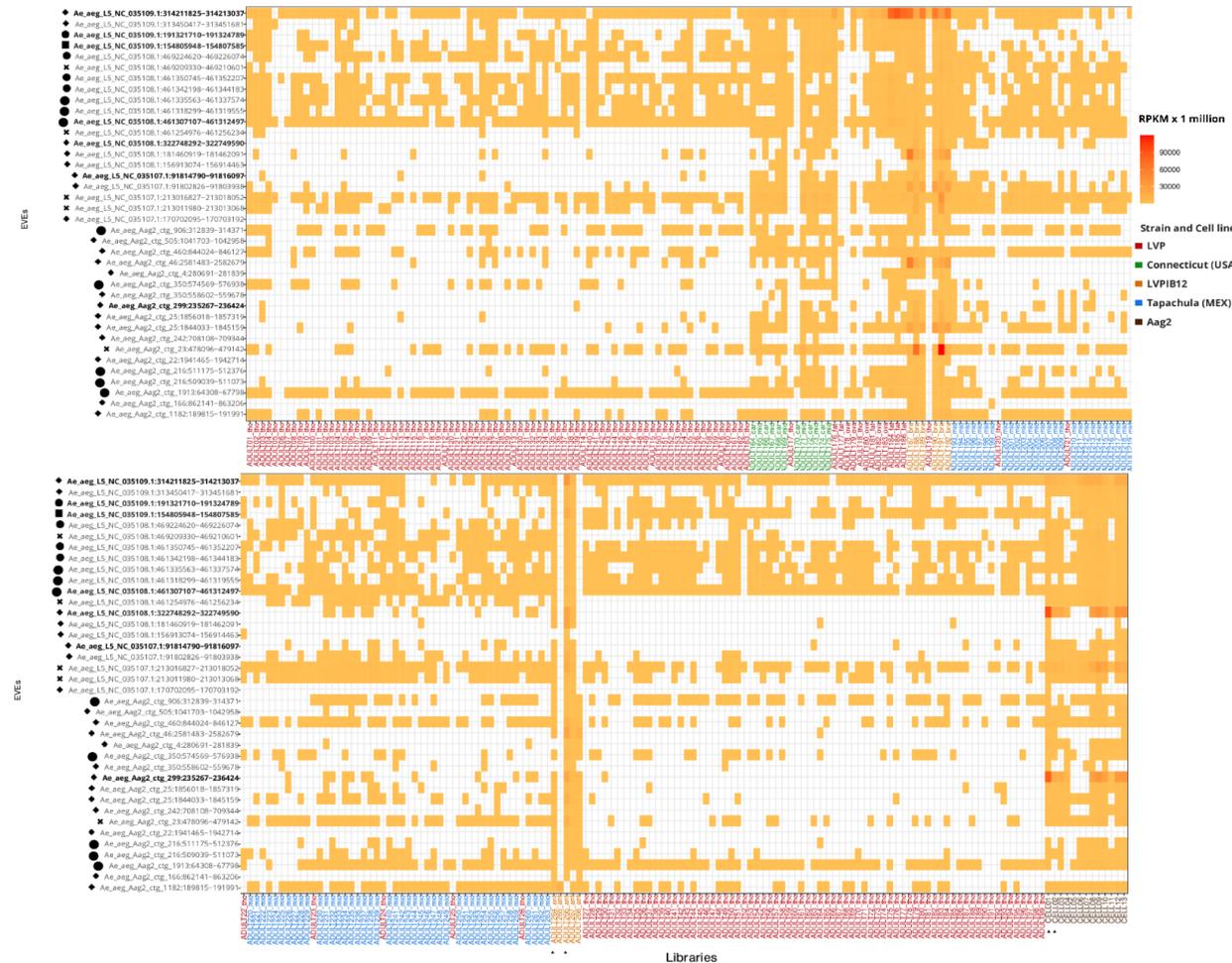
5.1 *Aedes aegypti*

Para essa espécie, o banco de dados do grupo de pesquisa apresentava EVEs das famílias Chuviridae, Rhabdoviridae, Flaviviridae e Xinmoviridae. Esses se expressaram em bibliotecas de RNA-seq de algumas colônias, de fase adulta e cultura celular. Para os de fase adulta, a pesquisa resultou em 174 bibliotecas da colônia Liverpool (LVP), obtidas por amostras de tórax, tecido adiposo ou ovário, 71 da colônia Tapachula (México), de intestino médio, 12 da Connecticut (USA), de tecido de carcaça ou intestino médio e 11 da colônia Liverpool IB12 (LVPIB12), com amostras de antena ou cérebro. Já para cultura celular, 13 bibliotecas foram encontradas, todas da linhagem Aag2.

O padrão da contagem de *reads* mapeados variou entre cada EVE. É possível ainda observar que, em determinadas bibliotecas, há ausência de *reads* mapeados. Alguns elementos demonstram-se promissores para análises *in vitro*. Para o EVE da família Chuviridae selecionado, observou-se que o mapeamento com a biblioteca ADULT 189 retornou um valor de RPKM de 3566,90. Já com a ADULT 192, foi obtido um total de 3926,70 RPKM. Os 2 EVEs selecionados da família Flaviviridae foram mapeados contra 2 diferentes bibliotecas. O 1º, com a ADULT 266, mapeou 6860,67 RPKM. Já o 2º EVE, com a ADULT 184, resultou em 8722,08 RPKM. As bibliotecas citadas são da linhagem LVPIB12.

Para a família Rhabdoviridae, 4 EVEs foram selecionados. O primeiro EVE, mapeado contra a biblioteca ADULT 264, retornou 7328,38 RPKM, contra a ADULT 266, 37577,56 RPKM e com a ADULT 267, 6659,24 RPKM, ambas de antenas de mosquitos adultos da colônia LVPIB12. O segundo, mapeado contra a ADULT 191, da mesma colônia, porém com amostras de tecido cerebral, em que foram expressos 16103,29 RPKM. Também houveram 2 EVEs da família cuja contagem de *reads* mapeados ocorreu em linhagens celulares. O primeiro, expresso na linhagem a CELL1, expressou 84630 RPKM e, no segundo caso, com a biblioteca CELL 2, foram mapeados 19789,68 RPKM. Não foram obtidos resultados satisfatórios para a família Xinmoviridae devido à baixa cobertura em todas as bibliotecas analisadas. Todos os resultados encontram-se resumidos na Figura 5.

Figura 5: Distribuição do padrão de contagem de *reads* mapeados, por espécime, tecido, fase evolutiva e colônia, associado aos EVEs, para *Ae. aegypti*.



Fonte: elaboradas pelos autores.

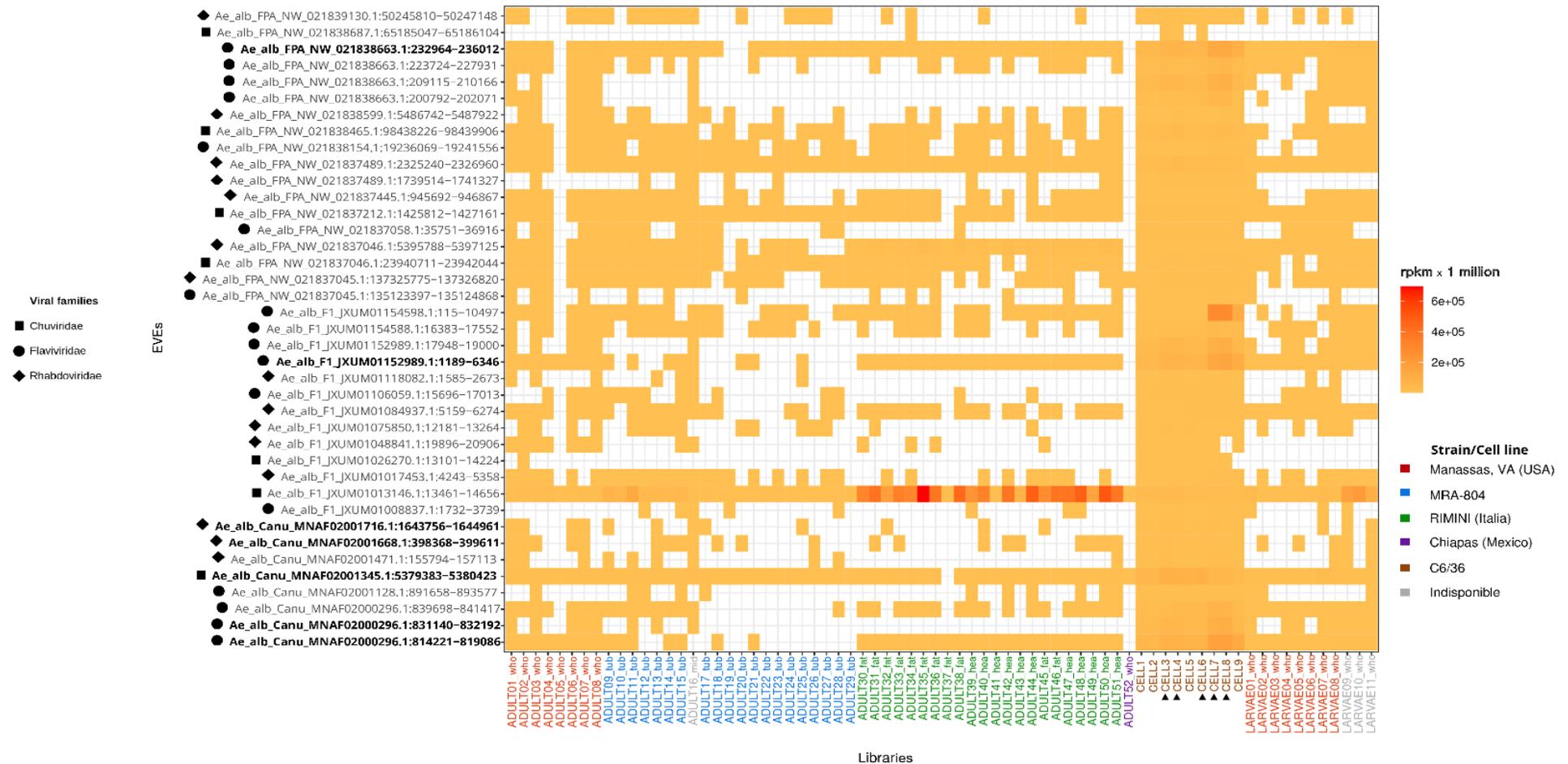
Legenda: mapa de calor representando a contagem de *reads* mapeados, por RPKM, através do cruzamento de dados entre bibliotecas de RNA-seq e EVEs. Espaços em branco representam a ausência de *reads* mapeados para a biblioteca. Tho = tórax; car = carcaça; mid = intestino médio; fat = tecido adiposo; ova = ovário; bra = cérebro; ant = antena. As setas (►) indicam as bibliotecas selecionadas para avaliação. Os EVEs selecionados estão em negrito.

5.2 AEADES ALBOPICTUS

Foram identificados nos dados de RNA-seq EVEs de três famílias virais: Flaviviridae, Chuviridae e Rhabdoviridae, distribuídos em diferentes linhagens e fases do desenvolvimento (Figura 6). 52 bibliotecas de RNA-seq foram utilizadas nos resultados, sendo 8 da colônia Manassas, VA (USA), representando amostras do corpo inteiro dos mosquitos, 19 da linhagem MRA-804, que utilizou os túbulos de Malpighi em sua composição, 21 da RIMINI (Itália), com amostras de cabeça e tecido adiposo, 1 da Chiapas (México), de corpo inteiro do mosquito e 1 sem colônia definida, originária do intestino médio desses insetos. 11 bibliotecas utilizaram mosquitos de fase larval em seus estudos. Destas, 8 eram da colônia Manassas, VA (USA) e 3 de cepa indefinida, ambas de corpo inteiro. Quanto a bibliotecas provenientes de estudos *in vitro*, 9 utilizaram cultura celular, todas da linhagem C636.

Apesar da grande quantidade de bibliotecas de RNA-seq de mosquitos adultos e larvas, os resultados mais promissores, devido à cobertura e profundidade de mapeamento, ocorreram para as amostras de cultura celular. Quanto à família Flaviviridae, quatro EVEs foram selecionados. O primeiro EVE que foi identificado na biblioteca, contra a biblioteca CELL 8, apresentou 131588,92 RPKM. O segundo mapeou 72747,72 RPKM contra a biblioteca CELL 4. O terceiro EVE da família, foi mapeado contra duas bibliotecas. Na CELL 7, foram retornados 93057,03 RPKM e na CELL 8, 104172,05 RPKM. O último EVE da família, contra a CELL 7 mapeou 163420,35 RPKM. Quanto à família Chuviridae, apenas um EVE foi selecionado. Este, mapeado contra a CELL 6 retornou 91636,53 RPKM e contra a CELL 3, 87585,58 RPKM. Por fim, para a família Rhabdoviridae, 2 EVEs foram selecionados para os estudos posteriores. O primeiro, contra a biblioteca CELL 3 mapeou 18820,75 RPKM. Já o segundo, retornou 18770,72 RPKM quando mapeado contra a biblioteca CELL 6. Todas as bibliotecas de cultura celular apresentadas utilizaram amostras da linhagem C636. Algumas das bibliotecas de adultos da linhagem RIMINI (Italia) mapearam uma quantidade de *reads* maior que todas as de cultura celular, no EVE da família Chuviridae, porém, a baixa cobertura desclassificou elas para os resultados. Ademais, os dados de adultos e larvas não atenderam aos critérios preestabelecidos. Os resultados encontram-se expostos na Figura 6.

Figura 6: Distribuição do padrão de contagem de *reads* mapeados, por espécime, tecido, fase evolutiva e colônia, associado aos EVEs correspondentes, para *Ae. albopictus*.



Fonte: elaborada pelos autores.

Legenda: mapa de calor representando a contagem de *reads* mapeados, por RPKM, por cruzamento de dados entre bibliotecas de RNA-seq e EVEs. Espaços em branco mostram a ausência de *reads* mapeados p/ a biblioteca. Who = Corpo inteiro; Tub = Túbulos de Malpighi; mid = intestino médio; fat = tecido adiposo; hea = cabeça. As setas (►) indicam as bibliotecas selecionadas para avaliação. Os EVEs selecionados estão em negrito.

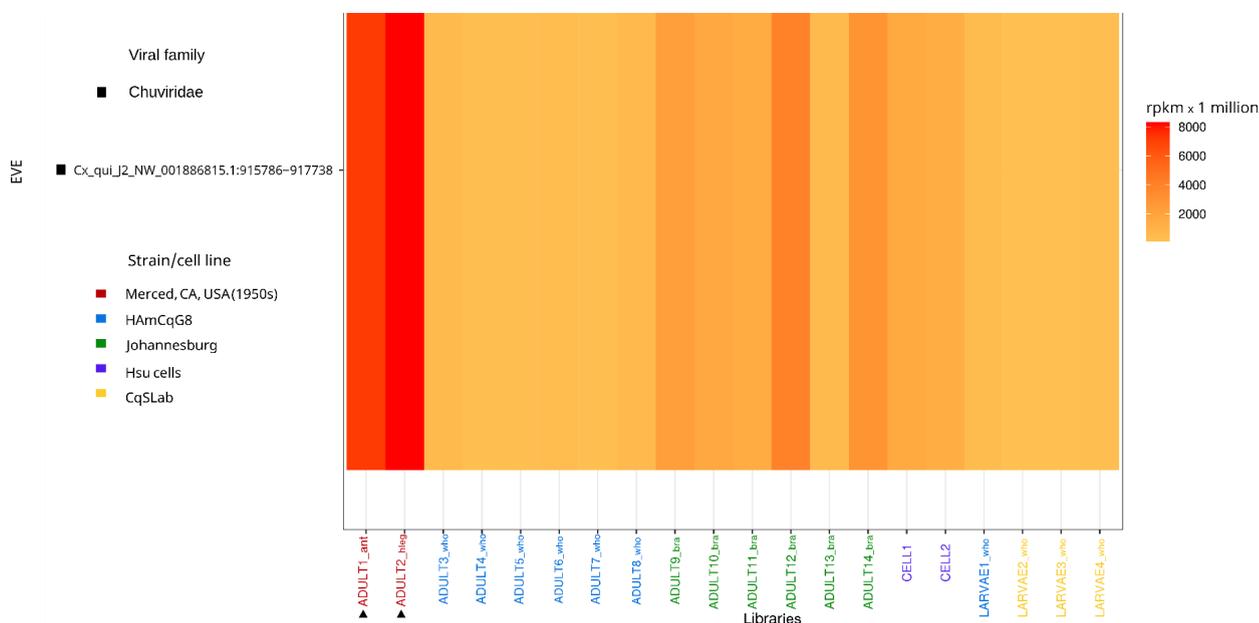
5.3 CULEX QUINQUEFASCIATUS

Para a espécie *Cx. quinquefasciatus*, apenas um EVE, da família Chuviridae, foi classificado para os resultados. Quanto à contagem de *reads* mapeados, é possível destacar 2 bibliotecas de adultos da colônia Merced, CA (1950s), com maior número de *reads* mapeados (Figura 7). Na biblioteca ADULT 1 (SRR991016), obtida da antena de mosquitos, foram mapeados 7200,82 RPKM. Já para a biblioteca ADULT 2, proveniente da perna traseira, o mapeamento resultou em 8309 RPKM.

Quanto à colônia HAmCqG8, todas as bibliotecas foram provenientes do corpo inteiro, tanto de adultos (ADULT 3-8) quanto de larva (LARVAE 1), e há uma evidente diferença quanto ao número de *reads* mapeados, comparada às bibliotecas anteriormente descritas. Nesse caso, o número variou entre 177,77 RPKM em ADULT 7 (SRR1005582) e 563,01 RPKM em ADULT 8 (SRR1005583), sendo ainda observado o valor de 401,64 RPKM para a LARVAE 1 (SRR1012913). Para a linhagem Johannesburg, com 6 bibliotecas de cérebro de adultos, o número de *reads* variou entre 512,81 RPKM, na ADULT 13 (SRR11292940), e 3983,61 RPKM, na ADULT 12 (SRR11292939).

Foram ainda definidas 2 bibliotecas de cultura celular, da linhagem Hsu cells. Foram mapeados 1643,95 RPKM da CELL 1 (SRR1539389) e 1346,31 RPKM da CELL 2 (SRR1539389). Quanto às larvas, além da LARVAE 1, também foram analisadas as bibliotecas LARVAE 2-4, da colônia CqsLab. Essas obtiveram os menores valores de RPKM, variando entre 118,34 RPKM na LARVAE 2 (SRR9617921) e 271,52 RPKM na LARVAE 3 (SRR9617922).

Figura 7: Distribuição do padrão de contagem de *reads* mapeados, por espécime, tecido, fase do desenvolvimento e colônia, associado ao EVE correspondente para *Culex quinquefasciatus*.



Fonte: elaborada pelos autores.

Legenda: mapa de calor representando a contagem de *reads* mapeados, por RPKM, através do cruzamento de dados entre bibliotecas de RNA-seq e EVEs. ant = antena; hleg = perna traseira; who = corpo inteiro; bra = cérebro. As setas (►) indicam as bibliotecas selecionadas para avaliação. Os EVEs selecionados estão em negrito.

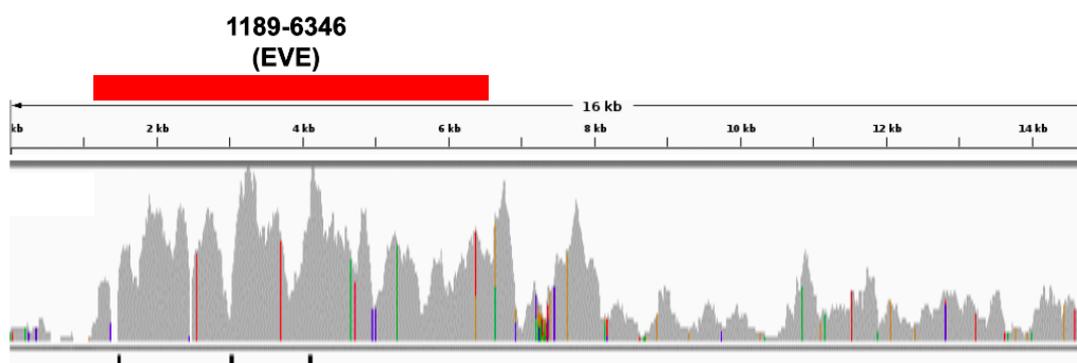
4.4 PROTEÍNAS

Diferentes domínios protéicos foram encontrados nos flancos dos EVEs de cada espécie analisada. A análise dos flancos de *Ae. aegypti* demonstrou a presença de integrases, transcriptases reversas, domínios de função desconhecida (DUF56 e DUF17), transposases, RdRP e nucleocapsídeos e glicoproteínas de rhabdovirus. Já na posição ocupada pelo EVE há as proteínas Flavi_NS1, DEAD-like e BamD.

Nos flancos dos EVEs de *Ae. albopictus*, os domínios protéicos presentes foram DUFs, peptidases, transcriptases reversas, AIR1 e integrases. Já as proteínas virais endogenizadas detectadas foram a Flavi_NS1, PT_UbiA, DEAD-like_helicase_C, P-look_NTPase, peptidase_S7, dentre outras. As proteínas, provenientes dos EVEs se relacionaram a diferentes padrões de profundidade no mapeamento, possivelmente associadas à transposons e mecanismos de piRNA

pode ter influenciado os padrões de profundidade mapeamento, sugerindo uma maior ou menor contagem de *reads* mapeados, comparadas às demais. É possível visualizar um exemplo em uma biblioteca de cultura celular, em que houve um aumento da profundidade do mapeamento, na região em que o EVE está disposto (Figura 9).

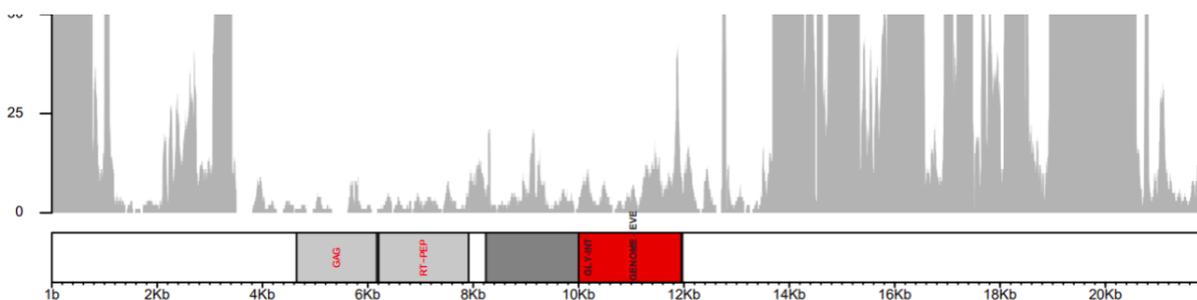
Figura 8: EVE da família Flaviviridae, associado à proteínas virais, modificando a profundidade de mapeamento.



Fonte: elaborada pelos autores.

Por fim, nos flancos de *Cx. quinquefasciatus*, foi detectada a presença de regiões associadas à peptidases, integrases, transcriptase reversas, domínios GAG e glicoproteínas, sendo a última também estendida até parte do EVE, podendo sugerir um processo de fusão. A presença de algumas destas proteínas pode ter influenciado os padrões de profundidade mapeamento. Um exemplo é observado em uma biblioteca de fase adulta, em que o EVE, associado à proteínas virais reduziu a profundidade do mapeamento (Figura 9). O resumo dos dados apresentados encontra-se no apêndice A.

Figura 9: EVE da família Chuviridae, associado à proteínas virais, modificando a profundidade de mapeamento.



Fonte: elaborada pelos autores. Legenda: Cor cinza: proteínas. Cor vermelha: EVEs. GAG: Domínio GAG. RT: Transcriptase reversa. PEP: Peptidase. GLY: Glicoproteína. INT: Integrase

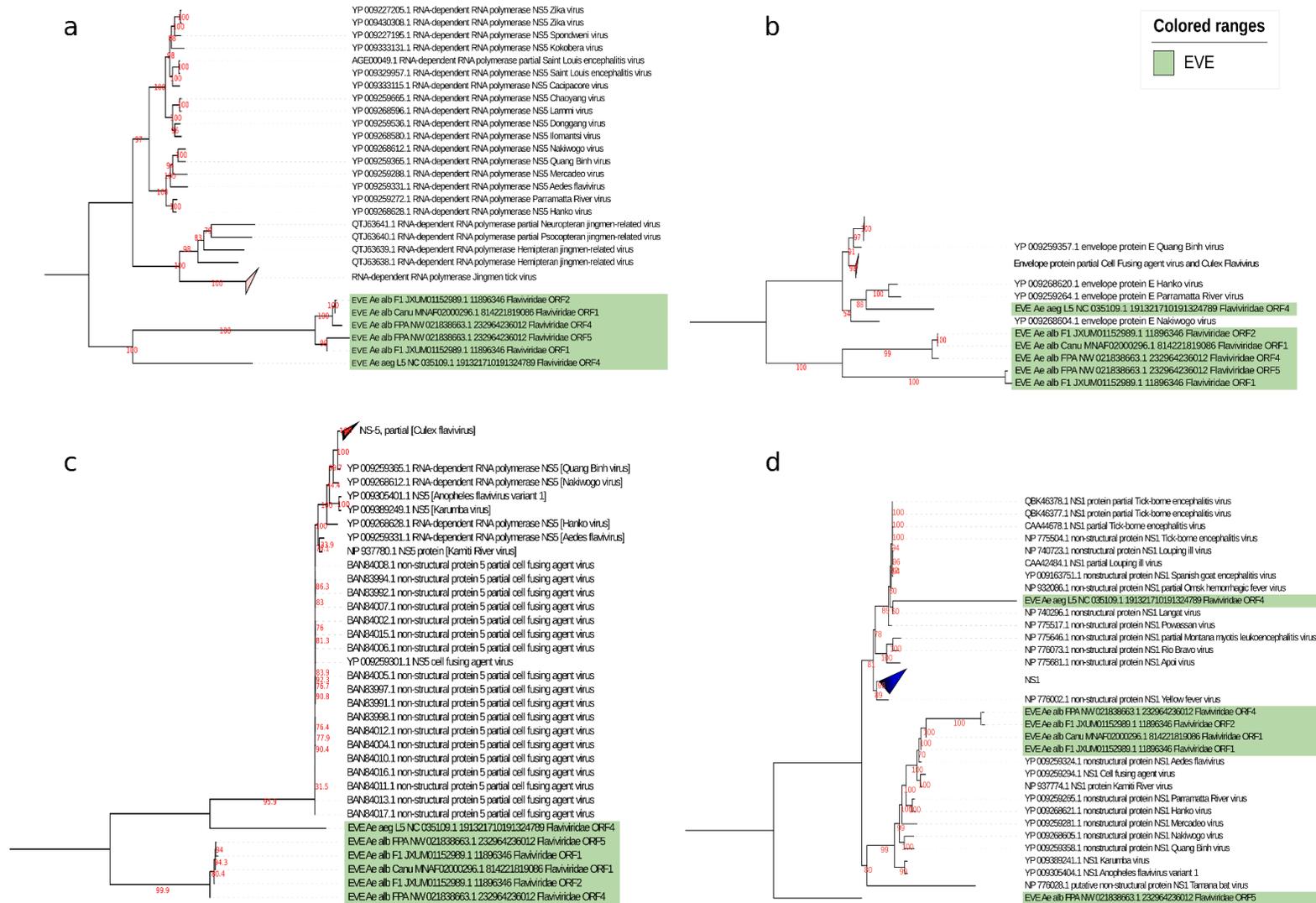
4.5 FILOGENIAS

Foi observada uma tendência em parte dos EVEs da família Flaviviridae a formar um clado basal, podendo inclusive ser ocupado por todos os elementos, como são os casos das filogenias associadas às proteínas RdRP (Figura 8.a) e NS5 (Figura 8.c). Para as proteínas de envelope, um evento semelhante ocorreu, porém um EVE foge ao padrão, agrupando-se com o Parramatta River virus e Hango virus (figura 8.b). Para a filogenia de NS1, houve uma maior dispersão das sequências por entre os clados. 4 EVEs formam um clado com a NS1 de *Aedes flavivirus*, além de 1 EVE basal e 1 agrupado juntamente com a NS1 do Langkat virus.

Para a família Rhabdoviridae, observa-se uma dispersão de EVEs em diferentes clados entre as nucleoproteínas. Um dos clados é composto por 4 EVEs de *Ae. aegypti* e algumas proteínas do nucleocapsídeo de diferentes vírus de *Ae. aegypti*, evidenciando a associação entre eles. 3 outros EVEs, no entanto, agrupam-se em um clado isolado, sem as demais sequências virais, demonstrando a individualidade destas quando comparadas às proteínas da família Rhabdoviridae (Figura 9).

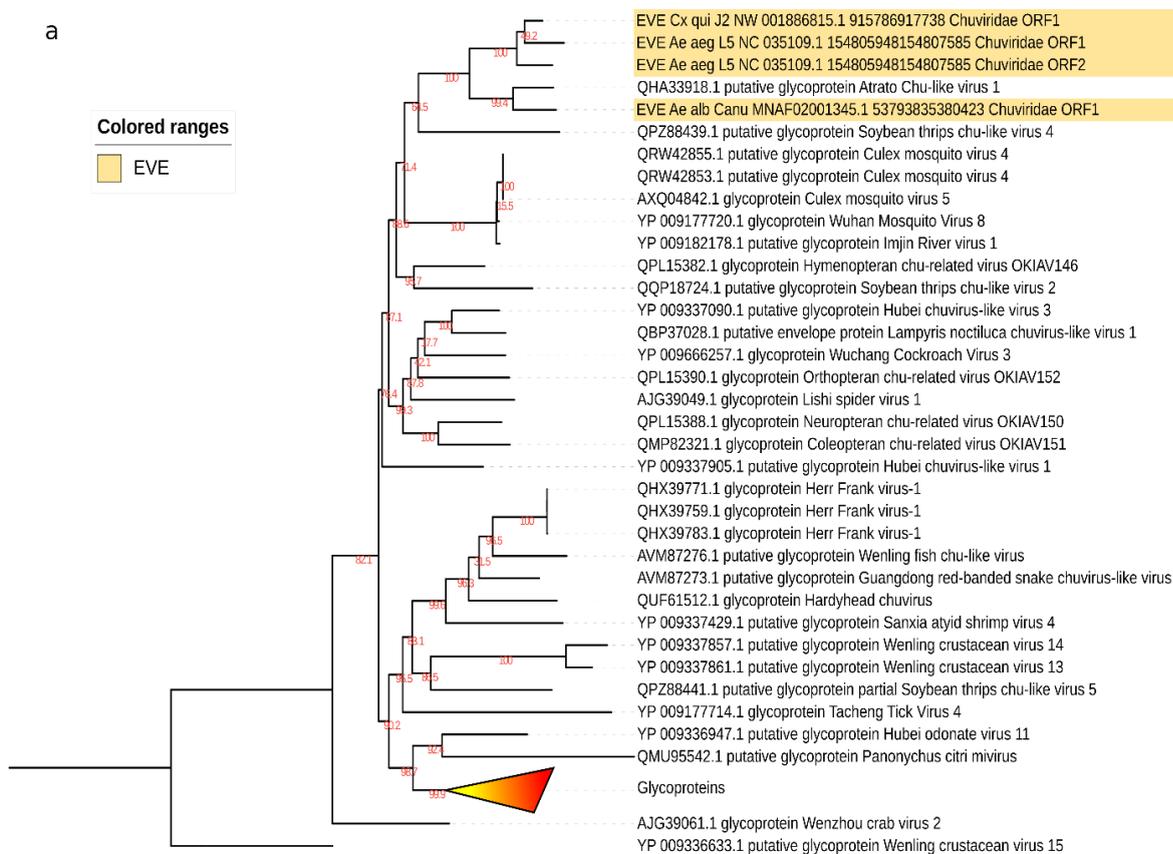
Para as glicoproteínas da família Chuviridae, foi também observada a formação de um agrupamento de EVEs em um clado. Porém, no mesmo clado em que esses se encontram, está também presente uma glicoproteína do Chu-like virus 1 (Figura 10.a), sugerindo que um vírus próximo pode ter dado origem ao EVE, possivelmente responsável por transferir o vírus endógeno ao genoma do hospedeiro eucarioto. Já para a filogenia de RdRP da mesma família, um clado basal é formado agrupando os EVEs com a proteína do Blattodean chu-related virus.

Figura 10: Filogenias associada aos EVEs da família viral Flaviviridae



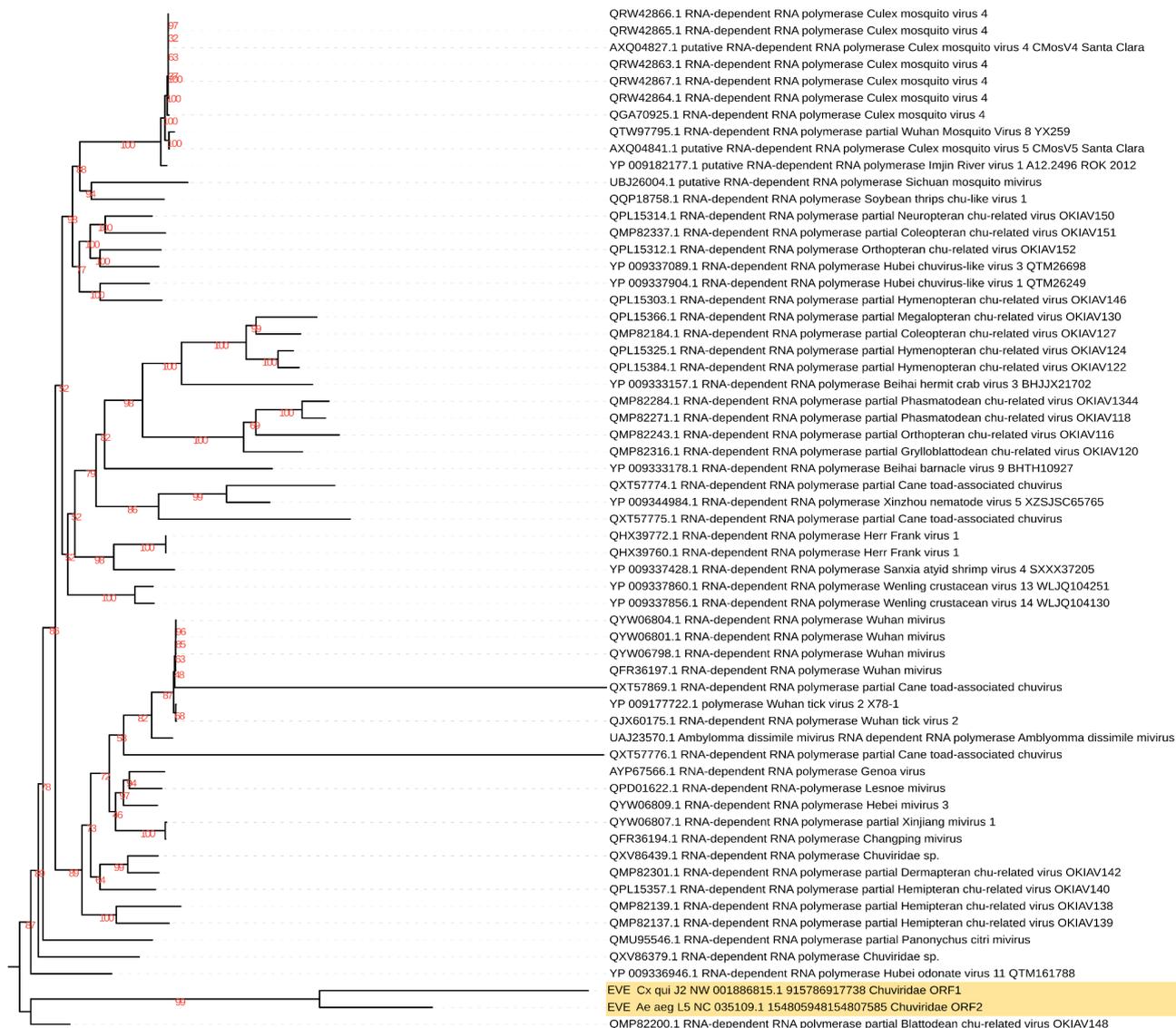
Fonte: elaborada pelos autores. a: Filogenia de RNA-polimerase dependente de RNA (RdRp); b: Filogenia de envelopes proteicos; c: Filogenia de proteínas não estruturais NS5; d: Filogenia de proteínas não estruturais NS1

Figura 12.a: Filogenia de glicoproteína associadas aos EVEs da família viral Chuviridae



Fonte: elaborada pelos autores.

Figura 12.b: Filogenia de RdRP associada aos EVEs da família viral Chuviridae



Fonte: elaborada pelos autores.

6 DISCUSSÃO

A nível celular, a modulação quantitativa de mRNAs é a via mais direta para a regulação da síntese proteica. Portanto, mecanismos que possam influenciar a expressão de mRNA podem também interferir na síntese de proteínas e consequentemente as funções biológicas daqueles organismos (HOERNES; HÜTTENHOFER; ERLACHER, 2016). Determinados culicídeos, são conhecidos por ter uma grande quantidade de EVEs em seus genomas (BLAIR; OLSON; BONIZZONI, 2020) e, portanto, é importante compreender as possíveis funções desses elementos na biologia do hospedeiro e sua influência em mecanismos regulatórios para mRNA.

De uma maneira geral, os EVEs que geraram maiores valores individuais de RPKM estavam associados à bibliotecas da espécie *Ae. albopictus*, seguida de *Ae. aegypti* e *Cx. quinquefasciatus*. Porém, a mediana do RPKM de cada espécie, considerando todas as bibliotecas encontradas foi de 225,51, 654,03 e 537,91, para *Ae. aegypti*, *Ae. albopictus* e *Cx. quinquefasciatus*.

Mesmo com o *Cx. quinquefasciatus* apresentando menor quantidade de EVEs, ele ainda apresentou uma mediana de RPKM maior que o *Ae. aegypti*, que possui uma grande quantidade de EVEs relatados na espécie (BLAIR; OLSON; BONIZZONI, 2020; RUSSO *et al.*, 2019). Surpreendentemente, os padrões de contagem de *reads* mapeados das duas espécies do gênero *Aedes* possuem uma diferença considerável, mesmo tratando-se de EVEs das mesmas famílias virais. Isso sugere que o número total de EVEs parece não ter associação direta com a contagem de *reads* mapeados, sendo algo inerente aos EVEs de maneira individual.

Há ainda diferenças entre padrões de cultura celular, principalmente no *Ae. albopictus*, que apresenta uma maior contagem de *reads* na linhagem C6/36. A diferença entre as culturas celulares analisadas pode ser uma hipótese que explicaria tais resultados, sendo necessários estudos de bancada, que avaliem, de modo comparativo, utilizando condições semelhantes, essas linhagens. Também observam-se diferenças nos resultados quando comparadas diferentes colônias e tecidos da mesma espécie.

Para *Cx. quinquefasciatus*, o número de *reads* expressos nos adultos da colônia Merced, CA, USA (1950s), em que foram utilizadas amostras de antena e perna traseira difere substancialmente das outras bibliotecas de mesma fase de

desenvolvimento, quando utilizados outras colônias, com amostras de tecidos diferentes ou de corpo inteiro. O mesmo é observado para *Ae. aegypti*, tratando-se da colônia LVPIB12, que mesmo em diferentes tecidos, é mais expresso a nível de RNA mensageiro que as colônias LVP, Connecticut (USA) e Tapachula (MEX), o que levanta hipóteses quanto a um processo evolutivo diferente entre as colônias. Isso também sugere esses fatores como possíveis interferentes quanto à contagem de *reads* mapeados, sendo mais uma vez necessário estudos complementares para compreender a real diferença entre os resultados.

Como demonstra Drongitis *et al.* (2019), existe uma possível associação entre elementos transponíveis (TEs) e a regulação gênica, que pode influenciar na estabilidade de mRNA por diferentes mecanismos. Entre eles estão a atividade de microRNAs, RNAs circulares e RNAs não codificantes potencialmente longos (lncRNAs). Como consequência, a expressão de proteínas traduzidas também pode ser alterada, através da inserção de TEs em regiões não traduzidas (URTs) de mRNAs. Como apresentado, diversos domínios proteicos foram encontrados em associação com as sequências apresentadas para cada espécie. Observando os karyoplots (apêndice b), com frequência há uma diferença notável na contagem de *reads* devido à presença de transposons.

Kitano *et al.* (2018) ainda demonstra uma possível atuação de determinados TEs na promoção ou inibição da tradução em um grande número de genes. Existe ainda a possibilidade de haver um processo de domesticação com a inserção de TEs em regiões codificadoras, podendo resultar na produção de quimeras e de TEs determinarem diferentes métodos de splicing de mRNA que podem implicar na codificação dos genes e surgimento de novos genes em processos evolutivos (DRONGITIS *et al.*, 2019). Visto isso, é importante avaliar como as proteínas resultantes de ORFs encontradas no trabalho podem estar associadas à atuação de transposons, impactando na promoção ou inibição da expressão de transcritos nas três espécies de mosquito.

A frequente tendência de EVEs a se agruparem em clados na região basal da filogenia pode ser explicada por diferentes fatores. É possível que EVEs agrupados em clados isolados representem EVEs que ganharam novas funções durante seu processo evolutivo, distinguindo-se das proteínas já existentes. Aquelas que agrupam-se com as demais proteínas podem representar inserções recentes, que

estejam associadas às funções daquelas proteínas presentes no mesmo clado. É possível ainda que essas proteínas virais dos vírus ancestrais ainda não tenham sido encontradas ou publicadas, que esses tenham sofrido grandes alterações no genoma, por mutações ou que descaracterizem-os dos vírus atuais. Há também a possibilidade que os ancestrais já tenham sido extintos, restando apenas as sequências de EVEs provenientes de hospedeiros já existentes naquela época.

7 CONCLUSÕES

Analisando os resultados *in silico*, foram selecionados 15 alvos principais para estudos funcionais *in vitro* que serão realizados posteriormente. Os resultados mostram importantes diferenças quando utilizadas bibliotecas de cultura celular, com diferentes linhagens, em que a contagem de *reads* pode ser maior, como nos resultados observados de *Ae. albopictus*, ou menor, como no *Cx. quinquefasciatus*. Foi também observada uma possível associação entre transposons na contagem de *reads* mapeados através do mecanismo de piRNAs, porém essa hipótese só pode ser confirmada através de estudos complementares. Ademais, as análises filogenéticas sugerem a existência de diferentes populações de EVEs nas 3 espécies de culicídeos e de eventos mutacionais, além de guiarem os resultados para determinadas espécies de vírus, em que os EVEs compartilharam maior similaridade.

REFERÊNCIAS

- AIEWSAKUN, Pakorn; KATZOURAKIS, Aris. Endogenous viruses: connecting recent and ancient viral evolution. **Virology**, v. 479-480, p. 26-37, maio 2015.
- ALMEIDA, Lorena Sampaio; COTA, Ana Lídia Soares; RODRIGUES, Diego Freitas. Saneamento, Arboviroses e Determinantes Ambientais: impactos na saúde urbana. **Ciência & Saúde Coletiva**, v. 25, p. 3857-3868, 2020.
- ARMEZZANI, Alessia *et al.* "Ménage à trois": the evolutionary interplay between JSRV, enjsrvs and domestic sheep. **Viruses**, v. 6, n. 12, p. 4926-4945, 9 dez. 2014.
- AYRES, Constância Flavia Junqueira *et al.* Zika virus detection, isolation and genome sequencing through Culicidae sampling during the epidemic in Vitória, Espírito Santo, Brazil. **Parasites & Vectors**, v. 12, n. 1, 8 maio 2019.
- BALTIMORE, David. Viral rna-dependent DNA polymerase: rna-dependent DNA polymerase in virions of RNA tumour viruses. **Nature**, v. 226, n. 5252, p. 1209-1211, jun. 1970.
- BEERNTSEN, Brenda T.; JAMES, Anthony A.; CHRISTENSEN, Bruce M. Genetics of mosquito vector competence. **Microbiology and Molecular Biology Reviews**, v. 64, n. 1, p. 115-137, 1 mar. 2000.
- BEIJERINCK, M. W. **Ueber ein Contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblätter**. Amsterdam: J. Müller, 1898.
- BENVENISTE, Raoul E.; TODARO, George J. Multiple divergent copies of endogenous C-type virogenes in mammalian cells. **Nature**, v. 252, n. 5479, p. 170-173, nov. 1974.
- BERMAN, Jules J. Viruses. **Taxonomic Guide to Infectious Diseases**, p. 263, 2019.
- BLAIR, Carol D.; OLSON, Ken E.; BONIZZONI, Mariangela. The widespread occurrence and potential biological roles of endogenous viral elements in insect genomes. **Current Issues in Molecular Biology**, p. 13-30, 2020.
- 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**, v. 100, n. 22, p. 13013-13018, 13 out. 2003.
- BOWTIE. **Bowtie 2**: fast and sensitive read alignment. 2018?. Disponível em: <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>. Acesso em: 6 jan. 2022.
- BRAGA, Ima Aparecida; VALLE, Denise. *Aedes aegypti*: histórico do controle no Brasil. **Epidemiologia e serviços de saúde**, v. 16, n. 2, p. 113-118, 2007.
- BRASIL, MINISTÉRIO DA SAÚDE; SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Guia de vigilância do *Culex quinquefasciatus***. 1. ed. Brasília, DF: Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância Epidemiológica, 2011.

BRASIL, Ministério Da Saúde. **Boletim Epidemiológico - Doenças tropicais negligenciadas 30 de janeiro – Dia mundial de combate às Doenças tropicais negligenciadas**. Brasília - DF: [s. n.], 2021.

BRONKHORST, Alfred W.; VAN RIJ, Ronald P. The long and short of antiviral defense: small RNA-based immunity in insects. **Current Opinion in Virology**, v. 7, p. 19-28, ago. 2014

CALLAHAN, R. *et al.* Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. **Proceedings of the National Academy of Sciences**, v. 79, n. 18, p. 5503-5507, 1 set. 1982.

CHAUHAN, Ravendra P.; GORDON, Michelle L. A systematic review analyzing the prevalence and circulation of influenza viruses in swine population worldwide. **Pathogens**, v. 9, n. 5, p. 355, 8 maio 2020.

CLAVERIE, J. Computational methods for the identification of genes in vertebrate genomic sequences. **Human Molecular Genetics**, v. 6, n. 10, p. 1735-1744, 1 set. 1997.

COFFIN, John.; HUGHES, Stephen; VARMUS, Harold. Retroviruses. **Cold Spring Harbor (NY)**: Cold Spring Harbor Laboratory Press, 1997.

COFFIN, John M *et al.* ICTV virus taxonomy profile: Retroviridae 2021. **Journal of General Virology**, v. 102, n. 12, 23 dez. 2021.

CONSORTIUM, Chimpanzee Sequencing and Analysis. Initial sequence of the chimpanzee genome and comparison with the human genome. **Nature**, v. 437, n. 7055, p. 69-87, set. 2005.

CONSORTIUM, International Chicken Genome Sequencing. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, v. 432, n. 7018, p. 695-716, 9 dez. 2004.

CONSORTIUM, International Human Genome Sequencing. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860-921, 15 fev. 2001.

CONSORTIUM, International Mouse Genome Sequencing. Initial sequencing and comparative analysis of the mouse genome. **Nature**, v. 420, n. 6915, p. 520-562, dez. 2002.

CROCHU, Sandrine *et al.* Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. **Journal of General Virology**, v. 85, n. 7, p. 1971-1980, 1 jul. 2004.

DEZORDI, Filipe Zimmer. Caracterização de vírus inseridos nos genomas de culicídeos vetores de patógenos. Recife: [s.n.], 2020.

DOS SANTOS, Taissa Pereira *et al.* Potential of *Aedes albopictus* as a bridge vector for enzootic pathogens at the urban-forest interface in Brazil. **Emerging Microbes & Infections**, v. 7, n. 1, p. 1-8, 28 nov. 2018.

DOUGHERTY, Robert M.; DI STEFANO, Henry S. Lack of relationship between infection with avian leukosis virus and the presence of COFAL antigen in chick embryos. **Virology**, v. 29, n. 4, p. 586-595, ago. 1966.

DRONGITIS, Denise *et al.* Roles of transposable elements in the different layers of gene expression regulation. **International Journal of Molecular Sciences**, v. 20, n. 22, p. 5755, 15 nov. 2019.

DUFFY, Siobain; SHACKELTON, Laura A.; HOLMES, Edward C. Rates of evolutionary change in viruses: patterns and determinants. **Nature Reviews Genetics**, v. 9, n. 4, p. 267-276, 4 mar. 2008.

DUNLAP, Kathrin *et al.* Endogenous retroviruses regulate periimplantation placental growth and differentiation. **Proceedings of the National Academy of Sciences**, v. 103, n. 39, p. 14390-14395, 2006a

DUNLAP, Kathrin; PALMARINI, Massimo; SPENCER, Thomas. Ovine Endogenous betaretroviruses (enJSRVs) and placental morphogenesis. **Placenta**, v. 27, p. 135-140, 2006b.

DUNWIDDIE, C. T. *et al.* Molecular cloning and characterization of gag-, pol-, and env-related gene sequences in the ev- chicken. **Journal of Virology**, v. 59, n. 3, p. 669-675, 1986.

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**, v. 102, n. 3, p. 725-730, 11 jan. 2005.

DUPRESSOIR, A. *et al.* Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. **Proceedings of the National Academy of Sciences**, v. 106, n. 29, p. 12127-12132, 29 jun. 2009.

FANG, Janet. Ecology: a world without mosquitoes. **Nature**, v. 466, n. 7305, p. 432-434, jul. 2010.

FERMIN, Damian *et al.* Novel gene and gene model detection using a whole genome open reading frame analysis in proteomics. **Genome Biology**, v. 7, n. 4, 2006.

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

FIOCRUZ. **Fiocruz identifica *Culex* no Recife com potencial para transmitir o vírus zika.** 21 jul. 2016. Disponível em: <https://agencia.fiocruz.br/fiocruz-identifica-culex-no-recife-com-potencial-para-transmitir-o-virus-zika>. Acesso em: 6 jan. 2022.

FORATTINI, Oswaldo Paulo. *Aedes (stegomyia) albopictus* (skuse) identification in brazil. **Revista de Saúde Pública**, v. 20, n. 3, p. 244-245, 1986.

FORATTINI, Oswaldo Paulo. Principais mosquitos de importância sanitária no Brasil. **Cadernos de Saúde Pública**, 1994.

FOSTER, Woodbridge A.; WALKER, Edward D. mosquitoes (Culicidae). *In: Medical and veterinary entomology*. Academic press, 2019. p. 261-325.

GARDNER, Lauren M.; CHEN, Nan; SARKAR, Sahotra. Global risk of Zika virus depends critically on vector status of *Aedes albopictus*. **The Lancet Infectious Diseases**, v. 16, n. 5, p. 522-523, maio 2016.

GEL, Bernat; SERRA, Eduard. KaryoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. **Bioinformatics**, v. 33, n. 19, p. 3088-3090, 29 maio 2017.

GEUKING, M. B. *et al.* Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. **Science**, v. 323, n. 5912, p. 393-396, 16 jan. 2009.

GUBLER, Duane. The emergence of epidemic dengue fever and dengue hemorrhagic fever in the Americas: a case of failed public health policy. **Revista Panamericana de Salud Pública**, v. 17, n. 4, abr. 2005.

GUBLER, Duane J. Dengue and dengue hemorrhagic fever. **Clinical Microbiology Reviews**, v. 11, n. 3, 18 dec. 2020.

GUBLER, Duane J. The economic burden of dengue. **The American journal of tropical medicine and hygiene**, v. 86, n. 5, p. 743, 2012.

GUEDES, Duschinka Rd *et al.* Zika virus replication in the mosquito *Culex quinquefasciatus* in Brazil. **Emerging Microbes & Infections**, v. 6, n. 1, p. 1-11, jan. 2017.

HANADA, Kousuke; SUZUKI, Yoshiyuki; GOJOBORI, Takashi. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. **Molecular biology and evolution**, v. 21, n. 6, p. 1074-1080, 2004.

HARDY, J. L. *et al.* Intrinsic factors affecting vector competence of mosquitoes for arboviruses. **Annual Review of Entomology**, v. 28, n. 1, p. 229-262, jan. 1983.

HÄRING, Monika *et al.* Independent virus development outside a host. **Nature**, v. 436, n. 7054, p. 1101-1102, ago. 2005

HAYWARD, A.; GRABHERR, M.; JERN, P. Broad-scale phylogenomics provides insights into retrovirus-host evolution. **Proceedings of the National Academy of Sciences**, v. 110, n. 50, p. 20146-20151, 25 nov. 2013.

HOERNES, Thomas Philipp; HÜTTENHOFER, Alexander; ERLACHER, Matthias David. mRNA modifications: dynamic regulators of gene expression? **RNA Biology**, v. 13, n. 9, p. 760-765, 28 jul. 2016.

HOLMES, Edward C. The evolution of endogenous viral elements. **Cell Host & Microbe**, v. 10, n. 4, p. 368-377, out. 2011.

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

JENKINS, Gareth M. *et al.* Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. **Journal of molecular evolution**, v. 54, n. 2, p. 156-165, 2002.

JOHNSON, Welkin E. Origins and evolutionary consequences of ancient endogenous retroviruses. **Nature Reviews Microbiology**, v. 17, n. 6, p. 355-370, 8 abr. 2019

KATZOURAKIS, Aris; GIFFORD, Robert J. Endogenous viral elements in animal genomes. **PLoS Genetics**, v. 6, n. 11, p. e1001191, 18 nov. 2010

KITANO, Shohei; KURASAWA, Hikaru; AIZAWA, Yasunori. Transposable elements shape the human proteome landscape via formation of cis-acting upstream open reading frames. **Genes to Cells**, v. 23, n. 4, p. 274-284, 2018.

KOONIN, Eugene V.; DOLJA, Valerian V.; KRUPOVIC, Mart. Origins and evolution of viruses of eukaryotes: The ultimate modularity. **Virology**, v. 479-480, p. 2-25, 2015

KOTSAKIOZI, Panayiota *et al.* Tracking the return of *Aedes aegypti* to Brazil, the major vector of the dengue, chikungunya and Zika viruses. **PLoS neglected tropical diseases**, v. 11, n. 7, p. e0005653, 2017.

KRZYWINSKI, M. *et al.* Circos: an information aesthetic for comparative genomics. **Genome Research**, v. 19, n. 9, p. 1639-1645, 18 jun. 2009.

KUMAR, Sudhir; SUBRAMANIAN, Sankar. Mutation rates in mammalian genomes. **Proceedings of the National Academy of Sciences**, v. 99, n. 2, p. 803-808, 2002.

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

LANGMEAD, Ben; SALZBERG, Steven L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357-359, 4 mar. 2012.

LANGMEAD, Ben *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, v. 10, n. 3, p. R25, 2009.

LARSSON, Anders. AliView: a fast and lightweight alignment viewer and editor for large datasets. **Bioinformatics**, v. 30, n. 22, p. 3276-3278, 5 ago. 2014.

LEINONEN, R.; SUGAWARA, H.; SHUMWAY, M. The sequence read archive. **Nucleic Acids Research**, v. 39, n. Database, p. D19—D21, 9 nov. 2010.

LERAT, E. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. **Heredity**, v. 104, n. 6, p. 520-533, 25 nov. 2009.

LETUNIC, Ivica; BORK, Peer. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. **Bioinformatics**, v. 23, n. 1, p. 127-128, 18 out. 2006.

- LETUNIC, Ivica; BORK, Peer. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. **Nucleic Acids Research**, v. 47, n. W1, p. W256—W259, 1 abr. 2019.
- LI, Heng et al. The sequence alignment/map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-2079, 2009.
- LIMA, Estelita *et al.* Insecticide resistance in *Aedes aegypti* populations from Ceará, Brazil. **Parasites & Vectors**, v. 4, n. 1, p. 5, 2011.
- LINDNER, Robert; FRIEDEL, Caroline C. A comprehensive evaluation of alignment algorithms in the context of rna-seq. **PLoS ONE**, v. 7, n. 12, p. e52403, 26 dez. 2012.
- LIMA-CAMARA, Tamara Nunes. Emerging arboviruses and public health challenges in Brazil. **Revista de Saúde Pública**, v. 50, 2016.
- LOWY, D. R. *et al.* Murine Leukemia Virus: High-Frequency Activation in vitro by 5-Iododeoxyuridine and 5-Bromodeoxyuridine. **Science**, v. 174, n. 4005, p. 155-156, 8 out. 1971.
- LÖWY, Ilana. Representing and intervening in public health: viruses, mosquitoes and Rockefeller Foundation experts in Brazil. **História, Ciências, Saúde-Manguinhos**, v. 5, n. 3, p. 647-677, 1999.
- KHAN, Arifa S. *et al.* Proposed algorithm to investigate latent and occult viruses in vaccine cell substrates by chemical induction. **Biologicals**, v. 37, n. 3, p. 196-201, 2009.
- MAGER, Dixie L.; STOYE, Jonathan P. Mammalian endogenous retroviruses. **Microbiology spectrum**, v. 3, n. 1, p. 3.1. 09, 2015.
- MARTIN, M. A. *et al.* Identification and cloning of endogenous retroviral sequences present in human DNA. **Proceedings of the National Academy of Sciences**, v. 78, n. 8, p. 4892-4896, 1 ago. 1981.
- MEDEIROS, Zulma *et al.* Controle da filariose linfática no Brasil, 1951 - 2000. **Epidemiol. Serv. Saúde**, Brasília, v. 12, n. 2, p. 77-86, jun. 2003.
- MEDSTRAND, P.; BLOMBERG, J. Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. **Journal of Virology**, v. 67, n. 11, p. 6778-6787, 1993.
- MI, Sha *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. **Nature**, v. 403, n. 6771, p. 785-789, fev. 2000.
- MIESEN, Pascal; JOOSTEN, Joep; VAN RIJ, Ronald P. PIWIs go viral: arbovirus-derived pirnas in vector mosquitoes. **PLOS Pathogens**, v. 12, n. 12, p. e1006017, 29 dez. 2016.
- MINH, Bui Quang *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. **Molecular Biology and Evolution**, v. 37, n. 5, p. 1530-1534, 3 fev. 2020.

MORAIS, Sirlei Antunes de; MARRELLI, Mauro Toledo; NATAL, Delsio. Aspectos da distribuição de *Culex (Culex) quinquefasciatus* say (diptera, culicidae) na região do rio Pinheiros, na cidade de São Paulo, estado de São Paulo, Brasil. **Revista Brasileira de Entomologia**, v. 50, n. 3, p. 413-418, set. 2006.

NCBI. **ORFfinder** - **NCBI**. 2004. Disponível em: <https://www.ncbi.nlm.nih.gov/orffinder/>. Acesso em: 6 jan. 2022.

OLSON, Ken E.; BLAIR, Carol D. Arbovirus–mosquito interactions: RNAi pathway. **Current Opinion in Virology**, v. 15, p. 119-126, dez. 2015

OLSON, Ken E.; BONIZZONI, Mariangela. Nonretroviral integrated RNA viruses in arthropod vectors: an occasional event or something more? **Current Opinion in Insect Science**, v. 22, p. 45-53, ago. 2017.

PAEZ-ESPINO, David *et al.* Uncovering Earth's virome. **Nature**, v. 536, n. 7617, p. 425-430, 17 ago. 2016

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

PANCETTI, Filipe Gabriel Menezes *et al.* Twenty-eight years of *Aedes albopictus* in Brazil: a rationale to maintain active entomological and epidemiological surveillance. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 48, n. 1, p. 87-89, fev. 2015.

PATEL, Sagar; PANCHAL, Hetalkumar; ANJARIA, Kalpesh. DNA sequence analysis by ORF FINDER & GENOMATIX tool: Bioinformatics analysis of some tree species of Leguminosae family. In: **2012 IEEE international conference on bioinformatics and biomedicine workshops**. IEEE, 2012. p. 922-926.

PAUPY, C. *et al.* *Aedes albopictus*, an arbovirus vector: from the darkness to the light. **Microbes and Infection**, v. 11, n. 14-15, p. 1177-1185, dez. 2009.

PAYNE, L. N.; CHUBB, R. C. Studies on the nature and genetic control of an antigen in normal chick embryos which reacts in the COFAL test. **Journal of General Virology**, v. 3, n. 3, p. 379-391, 1 dez. 1968.

POLISCHUK, Valery *et al.* Evidence for plant viruses in the region of Argentina Islands, Antarctica. **FEMS Microbiology Ecology**, v. 59, n. 2, p. 409-417, fev. 2007.

ROBINSON, James T. *et al.* Integrative genomics viewer. **Nature Biotechnology**, v. 29, n. 1, p. 24-26, jan. 2011.

ROY, S. *et al.* Regulation of reproductive processes in female mosquitoes. **Advances in Insect Physiology**, v. 51, p. 115-144, 2016.

RUSSO, Alice G. *et al.* Novel insights into endogenous RNA viral elements in *Ixodes scapularis* and other arbovirus vector genomes. **Virus Evolution**, v. 5, n. 1, 1 jan. 2019.

SALAZAR, Ma Isabel *et al.* Dengue virus type 2: replication and tropisms in orally infected *Aedes aegypti* mosquitoes. **BMC Microbiology**, v. 7, n. 1, p. 9, 2007.

SANJUÁN, Rafael. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. **PLoS Pathogens**, v. 8, n. 5, p. e1002685, 3 maio 2012.

SEVERSON, David; BEHURA, Susanta. Genome investigations of vector competence in *Aedes aegypti* to inform novel arbovirus disease control approaches. **Insects**, v. 7, n. 4, p. 58, 30 out. 2016.

SHIH, A.; MISRA, R.; RUSH, M. G. Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: relation to primate retroviruses. **Journal of Virology**, v. 63, n. 1, p. 64-75, 1989.

SIEBER, Patricia; PLATZER, Matthias; SCHUSTER, Stefan. The definition of open reading frame revisited. **Trends in Genetics**, v. 34, n. 3, p. 167-170, mar. 2018.

SUTTLE, Curtis A. Viruses in the sea. **Nature**, v. 437, n. 7057, p. 356-361, set. 2005.

TASSETTO, Michel *et al.* Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements. **eLife**, v. 8, 17 out. 2019.

TEMIN, Howard M.; MIZUTANI, Satoshi. Viral rna-dependent DNA polymerase: rna-dependent DNA polymerase in virions of rous sarcoma virus. **Nature**, v. 226, n. 5252, p. 1211-1213, 27 jun. 1970.

TER HORST, Anneliek M. *et al.* Endogenous viral elements are widespread in arthropod genomes and commonly give rise to piwi-interacting rnas. **Journal of Virology**, v. 93, n. 6, 19 dez. 2018.

THORVALDSDOTTIR, H.; ROBINSON, J. T.; MESIROV, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. **Briefings in Bioinformatics**, v. 14, n. 2, p. 178-192, 19 abr. 2012.

TRISTEM, M. *et al.* Characterization of a novel murine leukemia virus-related subgroup within mammals. **Journal of virology**, v. 70, n. 11, p. 8241-8246, 1996.

TORTORA, Gerard J.; CASE, Christine.; FUNKE, Berdell. Vírus, Viroides e Prions. *In*: TORTORA, Gerard J.; CASE, Christine.; FUNKE, Berdell. FOSTER, Woodbridge A.; WALKER, Edward D. mosquitoes (Culicidae). *In*: **Medical and veterinary entomology**. Academic press, 2019. p. 261-325.. Artmed Editora, 2016 .p.358- 388

VIEIRA, Genilton J.; FIOCRUZ. **Fiocruz identifica Culex no Recife com potencial para transmitir o vírus Zika**. 21 jul. 2016. Disponível em: <https://agencia.fiocruz.br/fiocruz-identifica-culex-no-recife-com-potencial-para-transmitir-o-virus-zika>. Acesso em: 7 out. 2021.

WEISS, Robin A. Exchange of genetic sequences between viruses and hosts. **Current topics in microbiology and immunology**, p. 1-29, 2017.

WEISS, Robin A. The discovery of endogenous retroviruses. **Retrovirology**, v. 3, n. 1, 3 out. 2006.

WHEELER, D. L. Database resources of the national center for biotechnology information. **Nucleic Acids Research**, v. 29, n. 1, p. 11-16, 1 jan. 2001.

WHITFIELD, Zachary J. *et al.* The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. **Current Biology**, v. 27, n. 22, p. 3511-3519.e7, nov. 2017.

WOODCROFT, Ben J.; BOYD, Joel A.; TYSON, Gene W. OrfM: a fast open reading frame predictor for metagenomic data. **Bioinformatics**, v. 32, n. 17, p. 2702-2703, 3 maio 2016.

YAP, Melvyn W. *et al.* Evolution of the retroviral restriction gene fv1: inhibition of non-mlv retroviruses. **PLoS Pathogens**, v. 10, n. 3, p. e1003968, 6 mar. 2014.

YIN, Tengfei; COOK, Dianne; LAWRENCE, Michael. Ggbio: an R package for extending the grammar of graphics for genomic data. **Genome Biology**, v. 13, n. 8, p. R77, 2012.

APÊNDICES

APÊNDICE A - DOMÍNIOS PROTEICOS PUTATIVOS CONSERVADOS

Quadro 1: Domínios proteicos putativos conservados encontrados para alinhamentos de proteínas das ORFs, enfatizando a região em que se encontram, para EVEs de *Aedes aegypti* (continua)

EVE	Bibliotecas de RNA-seq	Posição do EVE	Alinhamentos	Domínios proteicos putativos conservados	Posição da proteína
Ae_aeg_L5_NC_0351 09.1:191321710-191324789	SRR1167494	10000-13079	Putative gag polyprotein (QHA33694.1)	-	7065-5818
			Polyprotein (QIS62329.1)	-	22657-21428
			Polyprotein (QIS62329.1)	-	12102-10873
			AAEL007844-PA (EAT40425.1)	Rhabdo_glycop	17152-16010
			N_CAR (XP_021706382.1)	Gypsy	2262-1123
			polyprotein (QHA33695.1)	Integrase_H2C2	3985-2891
			HIP (HBS53814.1)	Transpos_IS481, rve, transpos_IS30	15329-14481
Ae_aeg_L5_NC_0351 108.1:461307107-461312497	SRR11232095	10000-15390	Glycoprotein (QPF16715.1)	DUF5641	2936-4615
			N_CAR (XP_035773911.1)	RT_like	89-1636
			HIP (OXA46575.1)	-	6084-7439
			N_CAR (XP_021703881.1)	DEAD-like_helicase_C	12111-13151
			polyprotein (QIS62329.1)	-	13191-14066
			AAEL000120-PA (EAT48851.1)	Rhabdo_ncap	18827-19690
			AAEL007866-PA (EAT40427.1)	-	21195-22046
Ae_aeg_L5_NC_0351 09.1:154805948-154807585	SRR11292945	10000-11637	N_CAR (XP_039430813.1)	-	11046-9934
	SRR11292922				
Ae_aeg_L5/NC_0351 09.1:314211825-314213037	SRR1167493	10000-11212	Transposase (KMQ82498.1)	-	14241-13423
	SRR11674934				
	SRR11674935				

Ae_aeg_L5_NC_0351 07.1:91814790-91816 097	SRR11292944	10000- 11307	N_CAR (XP_021694240.1)	BamD	11307-9967
			N_CAR (XP_029718849.1)	RdRP	12607-11852
Ae_aeg_Aag2_ctg_29 9:235267-236424	ERR1331745	10000- 11157	AAEL009940-PA (EAT38128.1)	Rhabdo_ ncap	9931-10656

Quadro 1: Domínios proteicos putativos conservados encontrados para alinhamentos de proteínas das ORFs, enfatizando a região em que se encontram, para EVEs de *Aedes aegypti* (conclusão)

EVE	Bibliotecas de RNA-seq	Posição do EVE	Alinhamentos	Domínios proteicos putativos conservados	Posição da proteína
Ae_aeg_L5_NC_0351 08.1:322748292-3227 49590	SRR10503152	10000- 11298	N_CAR (XP_021695295.1)	BamD e Rhabdo_Ncap	9931-11343

Legenda: N_CAR = Proteína não caracterizada; HIP = Proteína hipotética; CFAG - Vírus do agente de fusão celular

Fonte: os autores

Quadro 2: Domínios proteicos putativos conservados encontrados para alinhamentos de proteínas das ORFs, enfatizando a região em que se encontram, para EVEs de *Aedes albopictus* (continua)

EVE	Bibliotecas de RNA-seq	Posição do EVE	Alinhamentos	Domínios proteicos putativos conservados	Posição da proteína
Ae_alb_F1_JXUM 01152989.1:1189-6346	SRR8482205	1189-6346	N_CAR; HIP (XP_035773911.1; KXJ77309.1)	PH; DUF1759; RT_like; Peptidase; DUF56.	8470-16284
			CFAG polyprotein-like protein (AAQ03217.1)	PT_UbiA	5618-6397
			HIP (KXJ68152.1)	Peptidase_S7; P-loop_NTPase; DEAD-like_helicase_ C	3139-5676
			HIP (KXJ70526.1)	Flavi_NS1	1567-3039
Ae_alb_Canu_MN AF02000296.1:83 1140-832192	SRR8482215	10000-11052	HIP (KXJ77309.1)	PHD; DUF 1759; RT_Like; Peptidase e DUF564	81-6242
	SRR8482216		N_CAR (XP_035891812.1)	-	6426-7877
			HIP (KXJ70776.1)	DUF1759 e AIR1	12081-13451
			HIP (KXJ70776.1)	Pepsin_retropepsin_ 1; RT_like; Peptidase_A17; integrase e RVE	13935-17240
Ae_alb_Canu_MN AF02000296.1:81 4221-819086	SRR8482215	10000-14865	CFAG (AAQ03217.1)	Flavi_NS1; Peptidase; P-loop_NTP; DEAD-like	10.246-14916
			HIP (KXJ77309.1)	RT-Like; Peptidase_A; DUF	17000-23161
			N_CAR (XP_035891812.1)	-	23345-24796
Ae_alb_FPA_NW_ 021838663.1:232 964-236012	SRR8482216	10000-13048	N_CAR (XP_035901957.1)	RT_like; RNase_H_like; Integrase; rve; rve_3	7492-5126
			CFAG polyprotein-like protein (AAQ03217.1)	Peptidase_S7; P-loop_NTPase	11706-9862
			HIP (KXJ76460.1)	Pepsin_retropepsin_ like; RT_like	9028-7565
			HIP (KXJ68152.1)	P-loop_NTPase; DEAD-like_ helicase_C	4958-4110
			HIP (KXJ70526.1)	Flavi_NS1	12236-11397

Quadro 2: Domínios proteicos putativos conservados encontrados para alinhamentos de proteínas das ORFs, enfatizando a região em que se encontram, para EVEs de *Aedes albopictus* (conclusão)

EVE	Bibliotecas de RNA-seq	Posição do EVE	Alinhamentos	Domínios proteicos putativos conservados	Posição da proteína
Ae_alb_Canu_M NAF02001668.1:3 98368-399611	SRR8482207	10000- 11243	N_CAR (XP_021713297.1)	-	2137-872
			Putative RdRP (QPF16710.1)	RT_like	7791-6838
			CLUMA_CG00063 9; isoform A (CRK86808.1)	-	8997-8107
			Reverse transcriptase (AAX28844.1)	Rve	19316-18447
Ae_alb_Canu_MN AF02001716.1:16 43756-1644961	SRR8482204	10000- 11205	N_CAR (XP_029732085.1)	Rhabdo_ncap	11331-9997
			NS1-NS2AN-FIFO protein (YP_006491220.1)	-	14461-13721
			CFAG polyprotein-like protein (AAQ03217.1)	RT_like	15036-13543
			N_CAR (XP_035773911.1)	RT_like	15880-14945
			N_CAR (XP_035784663.1)	DUF1759	17043-15940
Ae_alb_Canu_M NAF02001345.1:5 379383-5380423	SRR8482207 SRR8482204	10000- 11040	Não foram encontrados domínios proteicos conservados		

Legenda: N_CAR = Proteína não caracterizada; HIP = Proteína hipotética; CFAG - Vírus do agente de fusão celular

Quadro 3: Domínios proteicos putativos conservados encontrados para proteínas resultantes de alinhamentos das proteínas traduzidas dos ORFs, enfatizando a região em que se encontram, para EVEs de *Culex quinquefasciatus*.

EVE	Bibliotecas de RNA-seq	Posição do EVE	Alinhamentos	Domínios proteicos putativos conservados	Posição da proteína
Cx_qui_J2_NW_0018868 15.1:915786-917738	SRR991016	10000- 11952	UNC (XP_035773911.1)	RT-Like e peptidase	6203-7912
			N_CAR (XP_035784663.1)	Gag domain	4646-6169
	SRR991017		N_CAR; Chuvirus Glycoprotein (XP_017464838.1; API61888.1)	Glycoproteína fusionada com integrase	8239-11970

Legenda: N_CAR = Proteína não caracterizada; HIP = Proteína hipotética;

Fonte: os autores

APÊNDICE B - MATERIAL SUPLEMENTAR

Estão disponibilizados, como material suplementar, no link <https://figshare.com/s/f94e35109b5eab187548> os mapas de calor, em alta qualidade, apresentados nas figuras 5, 6 e 7, karyoplots com os resultados do mapeamento de cada EVE e sua associação com domínios proteicos, e uma planilha com o todo o *dataset* utilizado para a pesquisa.