Universidade Federal de Pernambuco
Centro de Biociências

YAGO JOSÉ MARIZ DIAS

EEFINDER, UMA FERRAMENTA PARA A IDENTIFICAÇÃO DE ELEMENTOS ENDÓGENOS NO GENOMA DE EUCARIOTOS

Recife
2024

YAGO JOSÉ MARIZ DIAS

EEFINDER, UMA FERRAMENTA PARA A IDENTIFICAÇÃO DE ELEMENTOS ENDÓGENOS NO GENOMA DE EUCARIOTOS

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: Dr. Gabriel da Luz Wallau

Recife
2024

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

Mariz Dias, Yago José.
EEFINDER, UMA FERRAMENTA PARA A IDENTIFICAÇÃO DE ELEMENTOS ENDÓGENOS NO GENOMA DE EUCARIOTOS / Yago José Mariz Dias. - Recife, 2024.
49 p. : il., tab.

Orientador(a): Gabriel da Luz Wallau
Cooorientador(a): Filipe Zimmer Dezordi Zimmer Dezordi
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Biociências, Biomedicina, 2024.
Inclui referências, apêndices, anexos.

1. Elementos Virais Endógenos. 2. Elementos Bacterianos Endógenos. 3. Bioinformática. 4. Paleovirologia. I. da Luz Wallau, Gabriel. (Orientação). II. Zimmer Dezordi, Filipe Zimmer Dezordi. (Coorientação). IV. Título.

570 CDD (22.ed.)

YAGO JOSÉ MARIZ DIAS

EEFINDER, UMA FERRAMENTA PARA A IDENTIFICAÇÃO DE ELEMENTOS ENDÓGENOS NO GENOMA DE EUCARIOTOS

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: 04/03/2024

BANCA EXAMINADORA

_____

Orientador: Prof. Dr. Gabriel da Luz Wallau
Instituto Aggeu Magalhães - FIOCRUZ/PE / Departamento de Entomologia

_____

Dr. Alexandre Freitas da Silva
Instituto Aggeu Magalhães - FIOCRUZ/PE / Núcleo de Bioinformática

_____

Dr. João Luiz de Lemos Padilha Pitta
Instituto Aggeu Magalhães - FIOCRUZ/PE / Núcleo de Bioinformática

Dedico este trabalho ao meu avô,
José Bernardo Dias

AGRADECIMENTOS

Agradeço ao meu orientador Dr. Gabriel Wallau, por ter me aceito em seu grupo de pesquisa e com isso me inserido no mundo acadêmico de forma que conseguisse desenvolver habilidades que nunca antes imaginava adquirir. Ao meu co-orientador, Msc. Filipe Dezordi por reconhecer em mim o amor pela ciência e por se dispor a ser meu tutor. Sua presença constante ao meu lado, guiando, ensinando e alertando no desenvolvimento das minhas habilidades científicas, foi fundamental. Sem ele, todo o progresso até este momento teria sido significativamente mais difícil, senão impossível. Também quero agradecer a todos os colegas do meu laboratório que sempre estiveram prontos para ajudar, transformando nosso ambiente de trabalho em um espaço acolhedor e colaborativo. Sinto-me extremamente grato por fazer parte dessa equipe. Espero poder retribuir todo o apoio que recebi ao longo do caminho.

Quero agradecer à FACEPE por possibilitar a realização deste trabalho com uma bolsa de iniciação científica. Essa oportunidade tornou o processo mais acessível e proveitoso. Agradeço pelo suporte concedido.

Quero estender meus sinceros agradecimentos a todos os amigos, tanto aqueles que conheci durante a graduação quanto aqueles que deixei em minha cidade natal. A presença de vocês foi fundamental e tornou esse processo menos desgastante para mim. Em especial, gostaria de agradecer a Giovanna, que não só se tornou minha melhor amiga, mas também minha companheira, ajudando-me a superar traumas e construir independência emocional. Sua presença e apoio foram verdadeiramente transformadores.

Quero agradecer aos meus pais por terem depositado toda sua fé em um sonho, nunca medindo esforços para prover conforto e estrutura para o filho viver bem. Pai, agradeço por me ensinar a virtude da responsabilidade, nunca esquecerei o esforço que fez e ainda faz por mim. Mãe, agradeço por todos momentos que se preocupou com minha saúde emocional e sempre esteve ao meu lado quando pensei em desistir. Sem vocês, meu sonho seria apenas isso, um sonho. E por fim, quero agradecer ao meu avô José Bernardo Dias. Carregarei seu nome com honra e amor para sempre.

DIAS, Yago. **EEFINDER, UMA FERRAMENTA PARA A IDENTIFICAÇÃO DE ELEMENTOS ENDÓGENOS NO GENOMA DE EUCARIOTOS**. 2023. 49 folhas. Trabalho de Conclusão de Curso (Graduação em Biomedicina) – Universidade Federal de Pernambuco, Recife, 2024.

RESUMO

Transferência gênica horizontal (HGT) é um evento biológico onde ocorre a transferência de material genético entre espécies sem vínculo parental. Esse evento pode ocorrer entre espécies de diferentes níveis taxonômicos, desde gênero até reinos diferentes, como por exemplo entre bactérias/vírus e eucariotos. Os fragmentos de DNA derivados da HGT podem ser fixados na população caso a infecção ocorra em células da linhagem germinativa e dependendo da evolução da espécie pode ser fixado e transmitido para as próximas gerações. Esses elementos endogenizados (EEs) podem ser utilizados para inferir relações hospedeiro-patógeno de longo prazo. Além disso, esses elementos podem ser identificados como falsos positivos em análises de metagenômica viral, vigilância genômica e técnicas de identificação bacteriana por marcadores moleculares. Esses falsos positivos são gerados pela semelhança a nível genético com organismos circulantes, assim gerando falsos positivos nestes estudos. Esse problema pode ser resolvido com a montagem de um banco de dados com o conteúdo de EEs presentes no genoma de eucariotos. Atualmente não existem metodologias padronizadas para identificação desses elementos endógenos. Os estudos existentes usam metodologias diversas e filtragens manuais que aumentam o enviesamento dos resultados e dificultam a reprodutibilidade dos achados. Nesse contexto, uma ferramenta computacional padronizada se torna de extrema importância para aumentar a reprodutibilidade e avanço metodológico da identificação e caracterização dos elementos endógenos. Portanto o presente trabalho se propõe a desenvolver e testar uma ferramenta intitulada EEfinder que tem como objetivo identificar elementos endogenizados originados por HGT em genomas de eucariotos. A ferramenta foi desenvolvida em Python 3 utilizando o paradigma de programação orientada a objetos (POO). A execução do EEfinder apresenta 6 etapas: tratamento dos dados de entrada; busca dos elementos por algoritmos de alinhamento; filtragem dos elementos por banco de proteínas do hospedeiro; assinatura taxonômica; junção de elementos fragmentados e extração de regiões flanqueadoras. A ferramenta encontra-se disponível em https://github.com/WallauBioinfo/EEfinder, com documentação para instalação e instruções de uso. O teste de sensibilidade teve o objetivo de identificar EVEs (Elementos Virais Endogenizados) e EBEs (Elementos Bacterianos Endogenizados) demonstrando que a ferramenta atingiu resultados similares aos encontrados na literatura. Foi realizada a comparação entre os métodos de alinhamento (BLAST e DIAMOND), mostrando o BLAST como mais sensível. Nos testes de consumo de recursos computacionais, a ferramenta apresentou um baixo uso de recursos computacionais, podendo ser utilizado em computadores pessoais. O trabalho entrega uma ferramenta para identificação de elementos virais/bacterianos endógenos em genomas eucariotos que aumenta a reprodutibilidade no campo de pesquisa de elementos endógenos.

Palavras-chave: Bioinformática. Elementos Endógenos. Ferramenta. Genômica. Paleovirologia.

DIAS, Yago. EEFINDER, UMA FERRAMENTA PARA A IDENTIFICAÇÃO DE ELEMENTOS ENDÓGENOS NO GENOMA DE EUCARIOTOS. 2023. 49 folhas. Trabalho de Conclusão de Curso (Graduação em Biomedicina) – Universidade Federal de Pernambuco, Recife, 2024.

ABSTRACT

Horizontal gene transfer (HGT) is a biological event that occurs when species without a parental relationship exchange genetic material. This event can occur between species at different taxonomic levels, ranging from genera to different kingdoms, such as between bacteria/viruses to eukaryotes. DNA fragments derived from HGT may be fixed in the population when infection and HGT takes place at germline cells, and depending on the species evolution, it can then be fixed and transmitted to future generations. These endogenized elements (EEs) can be used to infer host-pathogen relationships and provide an advantage in studying ancient infections. Additionally, these elements can be identified as false positives in viral metagenomics, surveillance, and techniques for controlling culicidae, such as using mosquitoes infected with Wolbachia, due to their genetic similarity to circulating organisms, therefore generating false positives in these studies. This problem can be solved by assembling a database with the content of EEs present in the genomes of eukaryotes. However, currently there are no standardized methodologies for identifying these endogenous elements. Existing studies use diverse methodologies and manual filtering, which increase result bias and lower the reproducibility. In this context, a unified tool becomes extremely important to increase the reproducibility and methodological advancement of identifying and characterizing endogenous elements. Therefore, this study aims to develop and test a tool called EEfinder, which aims to identify endogenized elements originating from HGT in eukaryotic genomes. The tool was developed in Python 3 using the Object Oriented Programming (OOP) paradigm. EEfinder execution consists of 6 steps: data treatment; similarity analysis; element filtering by host protein database; taxonomic signature; merging fragmented elements and flank extraction. The tool documentation is currently available at https://github.com/WallauBioinfo/EEfinder. The sensibility tests had the objective to identify EVEs (Endogenous Viral Elements) and EBEs (Endogenous Bacterial Elements) showing similar results with the literature. A comparison was made between alignment modes (BLAST and DIAMOND), showing BLAST to be more sensitive, and computational requirements tests concluded low resource usage, making it suitable for personal computers. The study delivers a tool for identifying viral/bacterial endogenous elements in eukaryotic genomes that increases reproducibility in the field of endogenous element research.

Key words: Bioinformatics. Endogenous Elements. Tool. Genomics. Paleovirology.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

HGT	*Horizontal Gene Transfer*
EEs	Elementos Endogenizados
EVEs	Elementos Virais Endogenizados
EBEs	Elementos Bacterianos Endogenizados
ERVW-1	Gene da Sincitina-1
HERV-W	*Human endogenous retrovirus-W*
CHIKV	Vírus Chikungunya
WNV	*West Nile Virus*
DENV	Vírus da Dengue
ZIKV	Vírus Zika
YFV	*Yellow Fever Virus*
PCR	*Polymerase Chain Reaction*
NGS	*Next Generation Sequencing*
sRNA	*small RNA*
POO	Programação Orientada a Objeto
NCBI	*National Center for Biotechnology Information*

SUMÁRIO

6 CONCLUSÃO

REFERÊNCIAS

1 INTRODUÇÃO

Transferência horizontal gênica ou, do Inglês, *Horizontal gene transfer* (HGT), é um evento biológico que acontece quando existe a transferência de material genético entre indivíduos que não apresentam relação parental (Soucy; Huang; Gogarten, 2015). Esse evento já foi evidenciado em diferentes níveis taxonômicos, tanto em indivíduos de espécies diferentes quanto em níveis taxonômicos superiores, como reinos (Holmes, 2011). Tais eventos já foram descritos em *culicídeos,* o qual possui locus genômicos específicos que apresentam alta identidade com genomas virais (Whitfield *et al.*, 2017). Essas regiões com similaridade a genomas virais, são descritas como Elementos Virais Endogenizados (EVEs) (Katzourakis; Gifford, 2010). O mesmo pode ser notado com sequências homólogas a bactérias sendo encontradas em *loci* nos genomas de eucariotos (Leclercq *et al.*, 2016) gerando Elementos Bacterianos Endogenizados (EBEs).

Quando uma bactéria ou vírus infecta um eucarioto em suas células germinativas, por ocasião do próprio mecanismo de replicação (como nos retrovírus) ou por outro mecanismo, o material genético dessas entidades biológicas pode ser integrado no material genético do hospedeiro e a depender do sucesso evolutivo do hospedeiro, o elemento pode fixar-se em futuras populações (Aswad; Katzourakis, 2012). Os mecanismos para essas integração não estão completamente elucidados, mas tem-se como hipótese que elementos transponíveis, ou transposons, podem ser os responsáveis por essas integrações (Tassetto *et al.*, 2019). Transposons são elementos com capacidade de auto-replicação que residem nos genomas de eucariotos (McClintock, 1950). Por possuírem essa capacidade autônoma, alguns desses elementos (como retrotransposons) codificam proteínas que são capazes de reconhecer o RNA viral e inseri-los nos genomas de eucariotos (Wallau, 2022). Em bactérias esses eventos provavelmente ocorrem por união de extremidade não-homóloga quando o DNA bacteriano consegue ter contato com o núcleo eucarioto (Husnik; McCutcheon, 2018). Outros mecanismos de inserção incluem o próprio mecanismo replicativo de retrovírus que se integra ao DNA do hospedeiro para ser transcrito e traduzido (Holmes, 2011).

Os Elementos Endogenizados (EEs) podem ser utilizados para estudar relações antigas entre patógenos e hospedeiros (Souilmi *et al.*, 2021). Esses estudos apresentam grande desvantagem caso utilize-se apenas vírus/bactérias

atuais pois, as enzimas de replicação de bactérias e vírus apresentam taxas de erros maiores (Drake *et al.*, 1998; Sanjuán *et al.*, 2010), levando a perda do sinal mutacional desses patógenos, enviesando inferências feitas sobre infecções passadas (Katzourakis, 2013). Já as taxas mutacionais de eucariotos geralmente são menores que as desses patógenos (Drake *et al.*, 1998). Após a integração do vírus/bactéria esses elementos começarão a sofrer as taxas de mutação do hospedeiro (Aswad; Katzourakis, 2012). Isso permite estudos de infecções passadas mais precisos com esses elementos funcionando como fósseis virais (Katzourakis, 2013; Katzourakis; Gifford, 2010; Koutsovoulos *et al.*, 2014). Alguns desses elementos podem passar por um processo de domesticação no qual a proteína tem sua função modificada para ser utilizada em processos biológicos do hospedeiro (Aswad; Katzourakis, 2012). Um exemplo notável desse fenômeno é observado no contexto do gene humano *ERVW-1*, cuja função crucial envolve a fusão de membranas, desempenhando um papel vital no desenvolvimento placentário. Este gene é o produto de uma integração genômica e exibe uma identidade genética de 100% com o vírus endógeno HERV-W, conforme documentado por (Mi *et al.*, 2000). Este achado destaca a relevância do gene *ERVW-1* na biologia humana e ressalta sua origem a partir da incorporação evolutiva do vírus HERV-W. Esse processo também acontece em bactérias, como mostrado por Leclercq que descreveu uma inserção de 3 MB da bactéria *Wolbachia w*VulC no genoma do isópode *Armadillidium vulgare,* onde a integração desse elemento começou a determinar a frequência sexual do hospedeiro tornando-se o cromossomo sexual (Leclercq *et al.*, 2016).

Um problema gerado por essas integrações é a semelhança que esses elementos podem ter com patógenos circulantes (Dezordi *et al.*, 2020). Isso pode ocorrer quando o elemento é cooptado (quando um elemento que foi selecionado para uma função começa a ser selecionado e utilizado para uma função diferente) com o hospedeiro, aumentando a conservação daquela região (Aswad; Katzourakis, 2012). Estudos de metagenômica e vigilância podem identificar esses elementos como patógenos circulantes sendo considerados falsos positivos (Dezordi *et al.*, 2020; Lara Pinto *et al.*, 2017). Esse equívoco pode ser resolvido com um banco de dados constituído de elementos endógenos, onde estes elementos podem ser utilizados em etapas de filtro. Porém, atualmente, o processo de caracterização desses elementos apresenta uma baixa reprodutibilidade metodológica devido ao

uso de ferramentas diversas e a procedimentos manuais realizados por diferentes grupos de pesquisa (Palatini *et al.*, 2022).

A maioria dos estudos utilizam curagem manual dos elementos identificados, o que diminui a reprodutibilidade e aumenta os níveis de vieses desses resultados (Palatini *et al.*, 2022). A criação de uma ferramenta padronizada para caracterização de EVEs e EBEs torna-se um meio ideal para aumentar a reprodutibilidade e possibilitar comparações inter-estudo, possibilitando um refinamento metodológico necessário para descrição dos EEs em um genoma.

REVISÃO BIBLIOGRÁFICA

1.1 Doenças negligenciadas

Doenças negligenciadas são aquelas ligadas diretamente a locais com baixo acesso a saneamento básico, em contato com habitats de animais silvestres ou comércio desses animais, áreas com dificuldade de acesso e zonas de guerra (Engels; Zhou, 2020). Essas regiões em sua maioria encontram-se em zonas tropicais do continente Áfricano, Ásiatico e nas Américas. Essas doenças têm uma baixa iniciativa para erradicação pois não se encontram em grandes potências mundiais (Engels; Zhou, 2020). Além disso, não matam como outras doenças prioritárias, mas sim debilitam os atingidos, impedindo que trabalhem mantendo assim as populações no ciclo da pobreza (Da Conceição *et al.*, 2022; Engels; Zhou, 2020). Essas doenças, por se apresentarem majoritariamente em populações com baixo desenvolvimento socioeconômico, muitas vezes só são alvos de intervenções durante grandes surtos (Engels; Zhou, 2020). Mas também por terem uma baixa disseminação no globo não apresentam o apelo econômico para desenvolvimento de medicações e tratamentos (Engels; Zhou, 2020). O que mantém essas populações marginalizadas e reforçam estigmas de racismo e desigualdade social (Da Conceição *et al.*, 2022; Engels; Zhou, 2020).

Vários tipos de patógenos são responsáveis por essas doenças como: helmintos que induzem a esquistossomose, cisticercose; protozoários que induzem a leishmaniose, doença de Chagas; bactérias, que induzem por exemplo a tuberculose, hanseníase, sífilis; e vírus que causam a raiva, Dengue, Chikungunya e Zika. Muitos desses vírus são transmitidos por artrópodes hematófagos, como

carrapatos, flebotomíneos e em sua maioria mosquitos, sendo esses vírus reconhecidos como arbovírus (do inglês *arthropod-borne viruses*) (Young, 2018).

1.2 Arboviroses

A maioria dos arbovírus transmitidos por mosquitos são das famílias *Togaviridae*, *Bunyaviridae* e *Flaviviridae*, sendo a última de grande importância médica para humanos (Young, 2018). Dentre os principais arbovírus temos: o vírus Chikungunya (CHIKV), *West Nile Virus* (WNV), o vírus Dengue (DENV), o vírus Zika (ZIKV), o vírus da febre amarela, do Inglês, *Yellow Fever Virus* (YFV). Esses vírus apresentam um estágio replicativo no intestino de mosquitos que se alimentaram de um organismo infectado, a infecção então consegue alastrar-se por outros tecidos e atinge altos títulos nas glândulas salivares inoculando o patógeno em um novo organismo no próximo repasto sanguíneo (Young, 2018).

Existe uma expansão desses vírus para áreas antes não colonizadas como nos Estados Unidos com WNV e na europa com CHIKV e DENV (Barzon, 2018; Hadfield *et al.*, 2019). Muitos fatores podem ser apontados para a razão de tal expansão: crescimento populacional, maior variação de temperaturas, aumento da precipitação e aumento da resistência dos mosquitos, o que indica que nenhuma região está salva da colonização (Asad; Carpenter, 2018; Young, 2018).

Em 2022, de acordo com a Organização Pan-Americana da Saúde (OPAS), foram registrados mais de 3 milhões de casos de arboviroses (PAHO/WHO, 2023). Além dos mortos, muitos dos casos levam à sequelas com efeitos de longo prazo na população , como por exemplo, a infecção humana pelo vírus ZIKV relacionado com casos de microcefalia e síndrome de Guillain-Barré e a infecção pelo vírus CHIKV está atrelado com casos de poliartralgia (Weaver *et al.*, 2018). Além desses problemas, o impacto dessas doenças traz tanto problemas para o sistema de saúde e para a economia, diminuindo a capacidade de desenvolvimento do país (Young, 2018).

Os tratamentos para essas doenças, geralmente, são terapias de suporte, com exceções para tratamentos de imunoterapia passiva (Weaver *et al.*, 2018). Alguns medicamentos apresentam eficácia *in vitro* para alguns vírus (Barrows *et al.*,

2016; Weaver *et al.*, 2018). Existem vacinas sendo desenvolvidas para alguns arbovírus e temos exemplos de vírus que já apresentam vacinas disponíveis como YFV e DENV (Kallás *et al.*, 2024; Torres-Flores; Reyes-Sandoval; Salazar, 2022; Weaver *et al.*, 2018). Um dos principais métodos para controlar a infecção desses vírus é diminuir o contato dos humanos com os vetores, os mosquitos, controlando os locais de oviposição, utilizando mosquitos infectados com *Wolbachia*, uso de inseticidas e larvicidas (Wilson *et al.*, 2020).

1.3 Principais culicídeos vetores de arboviroses

A família Culicidae é composta atualmente por mais de 3000 espécies que estão distribuídas em duas subfamílias: Anophelinae, com 3 gêneros, e Culicinae com 110 gêneros (Mosquito Taxonomic Inventory, [*s. d.*]). Todas as espécies apresentam hábitos de repasto sanguíneo, necessários para o amadurecimento dos embriões, apenas o gênero *Toxorhynchites* não é considerado hematófago (Consoli; Oliveira, 1994).

No gênero *Culex,* temos mais de 800 espécies, sendo *Culex quinquefasciatus* e *Culex pipiens* os vetores mais estudados do gênero (Reis *et al.*, 2023). *Cx. quinquefasciatus* apresenta competência vetorial para ZIKV, DENV e CHIKV (Reis *et al.*, 2023). Já os mosquitos do gênero *Anopheles* são conhecidos por transmitir os patógenos causadores da malária e filariose, mas já foram detectados 35 arbovírus infectando esse gênero (Hernandez-Valencia *et al.*, 2023).

O gênero *Aedes,* apresenta dois grandes vetores de arbovírus: *Aedes aegypti* e *Aedes albopictus.* A primeira vez que o *Ae. aegypti* foi identificado como vetor foi em 1900 em Cuba (Souza-Neto; Powell; Bonizzoni, 2019). A competência vetorial é um fenótipo que depende de interações entre o mosquito, patógeno e a microbiota do mosquito (Souza-Neto; Powell; Bonizzoni, 2019). Por isso existem diferenças entre a competência vetorial de diferentes populações de *Ae. aegypti,* por exemplo as diferenças de competência vetorial entre as populações domésticas e silvestres de *Ae. aegypti* em uma mesma região (Carvalho-Leandro *et al.*, 2012). Mas em linhas gerais esse mosquito se apresenta como vetor primário para DENV, ZIKV, CHIKV e YFV (Souza-Neto; Powell; Bonizzoni, 2019).

O *Ae. albopictus* também se mostra como um importante alvo para o controle de vetores. Essa espécie que foi originada na Ásia, ocorre principalmente em áreas

rurais e suburbanas, com a globalização e mudanças climáticas esse mosquito foi introduzido nas Américas, África, Austrália, ilhas do Pacífico e Europa (Paupy *et al.*, 2009). Esse mosquito consegue transmitir os mesmo principais patógenos que o *Ae. aegypti* (os quatros sorotipos de DENV, CHIKV, ZIKV e YFV) apesar de não apresentar a mesma capacidade vetorial e ser muitas vezes descrito como vetor secundário (Lwande *et al.*, 2020). Uma grande diferença entre eles dois é a resistência do *Ae. albopictus* a condições climáticas desfavoráveis, tornando a distribuição do mosquito mais ampla (Briegel; Timmermann, 2001). Além disso, o mosquito realiza o repasto sanguíneo em um amplo número de espécies, possibilitando eventos de transbordamento de patógenos para outros hospedeiros (Benedict *et al.*, 2007).

1.4 Métodos de monitoramento viral

Uma das primeiras abordagens para monitorar a presença de vírus foi a maceração de tecidos e a inoculação em animais. Nesse método, seleciona-se um tecido de um animal ou mosquito suspeito de infecção, o qual é macerado. Em seguida, o material resultante é inoculado em um animal, permitindo a observação de sintomas típicos de infecção, como febre (Calisher; Maness, 1970). Pode-se também utilizar um animal sentinela, colocado em uma região estratégica ou utilizando os próprios animais daquele habitat e monitorando se o animal apresenta sintomas (Ramírez *et al.*, 2018). É possível ainda a inoculação em culturas de células, muito semelhante ao primeiro método, mas em escala menor, onde se pode até isolar o vírus (Leland; Ginocchio, 2007).

Atualmente as técnicas moleculares se sobressaem entre os outros métodos para estudos de vigilância e monitoramento por apresentarem uma alta sensibilidade e serem menos laboriosas (Ramírez *et al.*, 2018). Na técnica *Polymerase Chain Reaction* (PCR) pode-se utilizar *primers* específicos e detectar ácidos nucleicos virais até se não houver partículas virais viáveis (Ramírez *et al.*, 2018). Porém, tais técnicas necessitam de conhecimento da genômica viral, além dos custos de infraestrutura que a técnica necessita (Ramírez *et al.*, 2018).

O descobrimento viral também é de importância para o monitoramento, tanto para identificar novos vírus emergentes com potencial epidêmico, como para caracterizar os vírus presentes nesses vetores e entender a dinâmica do surgimento

dessas doenças (Maia, 2024). Advindo das tecnologias de sequenciamento de nova geração, do inglês *Next Generation Sequencing* (NGS), a metagenômica possibilita o sequenciamento de material genético viral presente em uma certa amostra (Zhang *et al.*, 2019). Estudos dessa área se mostram promissores, com descobertas de ordens virais inteiras antes desconhecidas (Zhang *et al.*, 2019).

1.5 Impacto de elementos endogenizados no monitoramento e controle

Os EVEs podem ser muito semelhantes aos vírus circulantes e isso pode gerar falsos positivos (Dezordi *et al.*, 2020; Nouri *et al.*, 2018). Técnicas que se baseiam na detecção de material genético podem sofrer com a interferência dos EVEs, os quais podem estar sendo transcritos e serem detectados em estudos metagenômicos como novos vírus (Dezordi *et al.*, 2020; Nouri *et al.*, 2018) assim interferindo nos resultados da vigilância viral.

O controle de culicídeos pode ser feito usando bactérias do gênero *Wolbachia*, existindo duas estratégias principais: liberar mosquitos machos infectados para gerar incompatibilidade citoplasmática nos gametas e impossibilitar o desenvolvimento da prole ou liberar mosquitos fêmeas infectadas com uma cepa específica de *Wolbachia* que diminui a replicação de arbovírus (Laven, 1967; Ye *et al.*, 2015). Mas para tais técnicas funcionarem precisa-se de um conhecimento prévio da disseminação de *Wolbachia* na população alvo (Yen; Failloux, 2020). Muitas vezes esse monitoramento da população do vetor e nível de infecção por *Wolbachia* pode ser feito por técnicas moleculares como a PCR convencional que detecta material genético da bactéria, porém não distingue se o material é de uma bactéria infectando o hospedeiro ou um EBE de *Wolbachia* integrado no hospedeiro (Inácio da Silva *et al.*, 2021).

1.6 Impacto biológico dos elementos endógenos

Além dos efeitos no monitoramento viral e controle de vetores, os EEs podem ser responsáveis por alguns efeitos biológicos no hospedeiro. Após a integração no genoma do hospedeiro o novo locus de EVE pode se fixar em uma região que transcreve *small RNA* (sRNA), podendo então contribuir ativamente para a modulação da replicação viral no hospedeiro (Ter Horst *et al.*, 2019). Outro

mecanismo imunológico descrito é baseado na transcrição e tradução de proteínas virais defectivas provenientes de EVEs (Nouri *et al.*, 2018). A presença de glicoproteínas com nível estrutural semelhante a circulantes que se ativamente traduzidas podem competir por sítio de ligação com vírus circulantes (Armezzani *et al.*, 2014; Nouri *et al.*, 2018). Além disso EVEs podem ter seus produtos proteicos utilizados em processos moleculares do hospedeiro, como acontece com o gene humano *ERVW-1* que transcreve uma proteína importante para fusão de membranas (Mi *et al.*, 2000). Os EBEs também foram descritos com esse tipo de comportamento, quando uma inserção completa de uma bactéria do gênero *Wolbachia* tomou a função do cromossomo W do isópode *Armadillidium vulgare,* agora determinando a frequência sexual do organismo (Leclercq *et al.*, 2016).

1.7 Métodos de estudo dos elementos endógenos

Identificações *in vitro* de EEs são possíveis, basta utilizar técnicas moleculares como a PCR e com *primers* específicos amplificar o material integrado (Leclercq *et al.*, 2016). Até mesmo utilizar sequências de EVEs conhecidos para recriar e observar partículas virais anciãs em laboratório para entender as interações dessas partículas e seus hosts *in vitro* (Katzourakis, 2013). Porém utilizar abordagens *in vitro* torna-se um processo custoso para identificar todo conteúdo endogenizado em um genoma, como por exemplo dos genomas de *Ae. aegypti* (AaegL5.0) com 1.3 GB (Matthews *et al.*, 2018) e *Ae. albopictus* (Aalbo_primary.1) com 2.5 GB (Matthews *et al.*, 2018). Por isso abordagens *in silico* tornam-se formas viáveis para identificação em massa desses elementos.

Os estudos *in silico* para identificação de EVEs geralmente seguem um fluxo similar, começando por uma filtragem por tamanho das sequências do genoma eucarioto para evitar comparações contra contaminações (Dezordi *et al.*, 2020; Palatini *et al.*, 2017). Para identificação em si utiliza-se uma ferramenta de alinhamento, geralmente BLASTx, da tradução do genoma do hospedeiro contra um banco de proteínas virais (Dezordi *et al.*, 2020; Whitfield *et al.*, 2017). Seguido de um filtro para retirar elementos putativos com similaridade a genes eucarióticos, que geralmente é realizado por um BLAST das regiões de EVEs putativos contra um banco de proteínas do hospedeiro (Palatini *et al.*, 2017; Whitfield *et al.*, 2017). Alguns estudos utilizam os flancos desses EVEs para pesquisas de transposons

interagindo com esses EVEs (Katzourakis; Gifford, 2010). Além disso, é comum realizar análises filogenéticas para entender as relações evolutivas que os EVEs identificados apresentam com os patógenos circulantes (Dezordi *et al.*, 2020; Whitfield *et al.*, 2017). Para EBEs, os estudos ainda continuam escassos e se utilizam de técnicas moleculares invés de bioinformática (Leclercq *et al.*, 2016).

A ferramenta mais utilizada na identificação EVEs é o BLASTx, pois comparações entre aminoácidos são mais sensíveis que entre nucleotídeos para detecção de homólogos distantes evolutivamente (Palatini *et al.*, 2022). Os estudos apresentam uma notável diferença entre o número de EVEs encontrados, mesmo utilizando o mesmo genoma, isso pode ocorrer por diversos motivos: algoritmos *in house* não documentados, bancos de proteínas virais diferentes, diferentes parâmetros e filtragens manuais (Whitfield *et al.*, 2017). Na curagem dos elementos muitas vezes não é descrito quais foram os critérios utilizados (Whitfield *et al.*, 2017). Estudos diferem em como anotar elementos em uma mesma região, sendo várias integrações de um mesmo vírus ou uma única integração (Palatini *et al.*, 2022). Dessa forma, o desenvolvimento de uma ferramenta se faz necessária para gerar resultados reprodutíveis e comparáveis (Palatini *et al.*, 2022).

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver uma ferramenta que possibilite a identificação de elementos endogenizados virais e/ou bacterianos em genomas eucariotos de forma reprodutível e automatizada.

2.1 Objetivos Específicos

- Desenvolver a ferramenta e disponibilizar o código de forma pública.
- Comparar os dois principais modos de alinhamento em larga escala disponibilizados na literatura.
- Analisar o consumo de recursos computacionais da ferramenta com testes com recursos limitados.
- Testar a sensibilidade da ferramenta para identificação de EVEs/EBEs comparando com resultados da literatura no genoma de eucariotos.

3 METODOLOGIA

3.1 Desenvolvimento da ferramenta

A ferramenta foi desenvolvida na linguagem de programação Python (Welcome to Python.org, 2024) versão 3.9. Essa linguagem foi escolhida por ser uma linguagem multiparadigma, facilidade no aprendizado e desenvolvimento por ser de alto nível, o que facilita o aprendizado da ferramenta, e por ser amplamente adotada na comunidade científica. Foram utilizadas abordagens para identificação de EEs já presentes na literatura como: filtro de sequências pequenas, filtro de elementos putativos por alinhamento com proteínas do hospedeiro e extração de flancos (Dezordi *et al.*, 2023; Palatini *et al.*, 2017; Russo *et al.*, 2019; Whitfield *et al.*, 2017). Utilizamos o paradigma de programação POO (Programação Orientada a Objeto) que promove modularidade, reusabilidade e manutenção do código, além da ferramenta ter sido transformada em pacote podendo ser integrada em novas aplicações. Durante o desenvolvimento utilizamos a ferramenta Git (Chacon; Straub, 2014) para versionamento e Github (github.com, [*s. d.*]) como repositório. O EEfinder funciona por linha de comando com uma série de parâmetros customizáveis (Tabela 1).

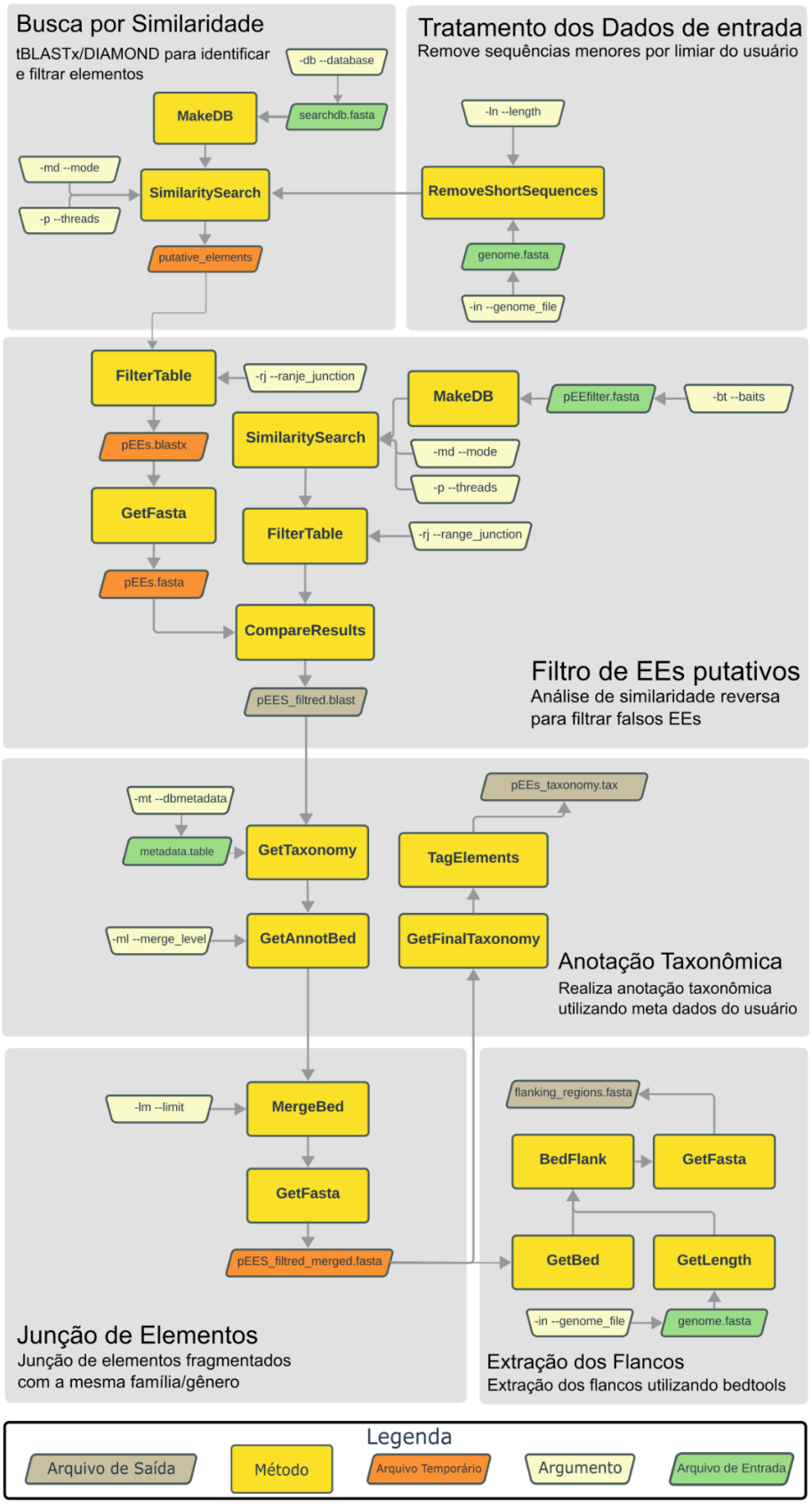
A ferramenta possui 6 etapas principais: tratamento dos dados; busca de similaridade; filtro de elementos putativos; assinatura taxonômica; junção de elementos fragmentados e extração dos flancos (Figura 1). O EEfinder requer 4 arquivos de entrada (Tabela 1 e Figura 1): o genoma do eucarioto alvo em formato fasta; o banco de dados proteico viral/bacteriano no formato fasta; uma tabela de metadados com os campos necessários para anotação taxonômica: ID de acesso, espécie, gênero, família, tipo de molécula, produto proteico, hospedeiro; e um arquivo *baits* em formato fasta com proteínas para serem utilizadas na filtragem de falsos positivos.

Tabela 1: Argumentos da ferramenta com valores esperados e breve explicação do uso.

Argumento	Valor esperado	Descrição
-in, --genome_file	Arquivo	Arquivo fasta do genoma eucarioto a ser pesquisado
-db, --database	Arquivo	Arquivo fasta com proteínas virais/bacterianas
-mt, --dbmetadata	Arquivo	Arquivo tabular com informações para anotação taxonômica
-bt, --baits	Arquivo	Arquivos com proteínas do hospedeiro para filtragem dos elementos putativos
-md, --mode	Texto	Seleção do algoritmo de alinhamento a ser utilizado, BLAST ou DIAMOND
-ln, --length	Número	Tamanho mínimo dos contigs do genoma
-fl, --flank	Número	Tamanho que será extraído dos flancos
-lm, --limit	Número	Limite para juntar elementos com a mesma taxonomia
-rj, --range_junction	Número	Intervalo de junção dos elementos redundantes
-mp, --mask_per	Número	Limite em porcentagem de caracteres minúsculos para considerar um EE uma região repetitiva
-cm, --clean_masked	Flag	Utilizar esse argumento remove EEs em regiões repetitivas
-p, --threads	Número	*Threads* utilizadas na análise
-rm, --removetmp	Flag	Utilizar esse argumento remove os arquivos intermediários da análise
-id, --index_databases	Flag	Realiza indexação dos bancos de dados
-pr, --prefix	Texto	Prefixo para os arquivos
-od, --outdir	Texto	Caminho onde serão guardados os arquivos de resultado
-ml, --merge_level	Texto	Nível que os elementos serão juntados

Fonte: Autor, 2024

Figura 1: Fluxo computacional da ferramenta



Fonte: Autor, 2024

3.1.1 Tratamento dos Dados

EEfinder foi desenvolvido usando as metodologias utilizadas por diferentes estudos já publicados. Por exemplo, alguns estudos usam um limiar para o tamanho das sequências do genoma pesquisado onde os EEs foram encontrados (Flynn; Moreau, 2019; Palatini *et al.*, 2017), filtrando possíveis contaminações que poderiam ser identificadas como EE. Logo, o EEfinder permite a seleção do tamanho mínimo de *contigs/scaffolds* que são utilizados na etapa de busca (Figura 1).

3.1.2 Busca por Similaridade

O EEfinder foi desenvolvido com dois principais algoritmos para realizar os alinhamentos, escolhidos pelo usuário, BLAST v2.5.0 (Camacho *et al.*, 2009) ou DIAMOND v2.0.15 (Buchfink; Reuter; Drost, 2021) (Figura 1). Utilizando os dois algoritmos, podemos encontrar elementos endógenos bastante divergentes dos patógenos circulantes (Palatini *et al.*, 2022). Comparações entre aminoácidos são mais sensíveis que com sequências nucleotídicas para detecção de elementos divergentes dos vírus atuais utilizados como sonda (Palatini *et al.*, 2022). No modo BLAST utilizamos o parâmetro word_size 3, que não está disponível no DIAMOND, e a matriz de substituição *BLOSUM45* para aumentar a sensibilidade da ferramenta.

3.1.3 Filtro dos EEs Putativos

Os resultados da etapa anterior são passados por um filtro para identificar falsos positivos. Sequências identificadas como elementos endógenos podem por acaso apresentarem uma convergência evolutiva com eucariotos (Dezordi *et al.*, 2020; Whitfield *et al.*, 2017). Esse problema é resolvido atualmente realizando filtragens manuais retirando falsos positivos (Dezordi *et al.*, 2020; Whitfield *et al.*, 2017), diminuindo a reprodutibilidade do método. Por esse motivo, o EEfinder aplica essa filtragem de forma automática realizando uma análise de similaridade dos elementos putativos contra um banco de proteínas do hospedeiro, contido no arquivo *baits* (Figura 1). Com os resultados desse alinhamento o EEfinder compara os *bitscores* dos resultados e mantém apenas elementos putativos que tiveram um

valor de *bitscore* maior nos resultados contra o banco viral/bacteriano.

3.1.4 Anotação Taxonômica

O EEfinder utiliza uma tabela de metadados provida pelo usuário para construir a assinatura taxonômica de cada EE (Figura 1). A ferramenta conta com instruções para montagem da tabela de metadados, também é disponibilizado um *script* adicional para construir essa tabela de forma automatizada com dados do usuário.

Essas classificações serão intermediárias pois serão utilizadas para a próxima função, de junção de elementos fragmentados, e após a função finalizada, é feita uma atualização da taxonomia intermediária para o resultado final.

3.1.5 Junção de Elementos Fragmentados

Elementos fragmentados podem ocorrer por acúmulo de mutações do hospedeiro, situação essa demonstrada por elementos ancestrais que durante as gerações do hospedeiro podem ter sido fragmentados (Aswad; Katzourakis, 2012) ou terem sido integrados de partículas virais defeituosas (Palatini *et al.*, 2022) e agora se apresentam como um elemento degenerado. A ferramentas juntará elementos que estiverem a pelo menos 100 nucleotídeos (valor padrão, que pode ser alterado pelo usuário) que representem a mesma família/gênero de vírus ou bactéria (Figura 1). Esse valor foi escolhido arbitrariamente visto que na literatura não existe essa abordagem.

3.1.6 Extração dos Flancos

Por fim, o EEfinder realiza uma extração dos flancos de cada elemento endógeno com uma opção de escolha do tamanho da extração (Figura 1). Esses flancos podem ser utilizados em análises adicionais em busca de elementos transponíveis ou identificação de genes do hospedeiro (Dezordi *et al.*, 2023; Whitfield *et al.*, 2017).

3.2 Dados de Teste

Nos testes de sensibilidade para EVEs utilizamos a versão do genoma Aag2 do espécie *Ae. aegypti* disponível com o código GCA_021653915 (https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/datasets/genome/) no *National Center for Biotechnology Information* (NCBI). Para o banco de proteínas virais utilizamos duas alternativas: todas proteínas do NCBI virus refseq na data 09/08/2022 e proteínas de EVEs descritas por Whitfield em seu estudo, que foi escolhido por ter o maior número de etapas automatizadas presente na literatura (Whitfield *et al.*, 2017). Atualizamos a taxonomia do estudo de Whitfield realizando um BLASTx contra o NCBI virus refseq no dia 11/10/2022. O banco de proteínas do hospedeiro foi obtido recuperando todas proteínas de *Ae. aegypti* no refseq na data 09/08/2022, duas versões foram feitas desse banco: proteínas não caracterizadas (sem filtrar as proteínas *hypothetical* e *uncharacterized*) e proteínas caracterizadas (filtrando o banco anterior retirando todas proteínas com termos *hypothetical* e *uncharacterized)*.

Para avaliar a sensibilidade para a detecção de EBEs e o consumo de recursos computacionais utilizamos o genoma do isópode *Armadillidium vulgare* GCA_001887335.1 contra um banco de proteínas da bactéria *W*vulC recuperado do NCBI na data 27/05/23. Para a filtragem dos resultados utilizamos todas proteínas do *Armadillidium vulgare* presentes no NCBI refseq na data 30/05/2023.

Tabela 2: Arquivos utilizados em cada análise com descrição do conteúdo

Arquivo	Conteúdo	Uso
Aag2.fa	Genoma de *Aedes aegypti*	Análise das diferenças na eficiência entre os modos de busca; Consumo de recursos computacionais; Análise de sensibilidade para EVEs
host_sub_characterized_protein.fa	Proteínas caracterizadas de *Aedes aegypti* (30/05/2023)	Análise das diferenças na eficiência entre os modos de busca; Consumo de recursos computacionais; Análise de sensibilidade para EVEs
refferences_whitfield.fasta	Proteínas descritas por Whitfield	Análise das diferenças na eficiência entre os modos de busca; Consumo de recursos computacionais; Análise de sensibilidade para EVEs
A_vulgare_v1.fna	Genoma de *Armadillidium vulgare*	Consumo de recursos computacionais; Análise de sensibilidade para EBEs
wvulc.fa	Proteínas da bactéria WvulC (27/05/23)	Consumo de recursos computacionais; Análise de sensibilidade para EBEs
armadillium_characterized_host_db.fasta	Proteínas caracterizadas de *Armadillidium vulgare* (30/05/2023)	Consumo de recursos computacionais; Análise de sensibilidade para EBEs

Fonte: Autor, 2024

3.3 Análise das diferenças na eficiência entre os modos de busca

Para determinar a indicação de uso para cada modo de busca (BLAST x DIAMOND), realizamos testes com a ferramenta utilizando os dois modos, com 8 *threads* em todas análises. Utilizando o BLASTx e DIAMOND blastx nos modos *fast, sensitive, mid-sensitive, more-sensitive, very-sensitive* e *ultra-sensitive* (Buchfink; Reuter; Drost, 2021). Após as análises comparamos os resultados em número de EVEs, intervalo de identidade dos EVEs com os vírus circulantes e tempo de execução da ferramenta.

3.4 Consumo de recursos computacionais

Para estipular o consumo de recursos computacionais da ferramenta, nós realizamos testes com números limitados de memória RAM e *threads* para bancos de dados virais e bacterianos. Os testes foram realizados com 4 configurações de memória RAM e threads: 4 *threads* com 8 GB, 8 *threads* com 16 GB, 16 *threads* com 32 GB, e 32 *threads* com 64 GB. Os testes foram repetidos três vezes para cada configuração.

3.5 Análise de sensibilidade para EVEs

Para avaliar a capacidade da ferramenta em identificar EVEs, configuramos a ferramenta com os mesmo parâmetros utilizados e com o mesmo conjunto de proteínas virais que Whitfield (Whitfield *et al.*, 2017) utilizou para descrever o conteúdo de EVEs no genoma Aag2 utilizando o seguinte comando:

eefinder -in Aag2.fa -mt refferences_whitfield.fasta.meta -db refferences_whitfield.fasta -bt host_sub_characterized_protein.fa -p 36 -ml family -lm 100

Para verificar se as diferenças numéricas entre o estudo de Whitfield e o EEfinder, geradas pela função de junção de elementos fragmentados, utilizamos *scripts in house*, que analisavam se a posição de começo e final de cada elemento

estavam sobrepostos com no máximo 100 nucleotídeos de diferença, para assim validar se os elementos dos dois estudos estavam em sobreposição.

A fim de concluir se Whitfield *et al* identificou um elemento endogenizado degenerado como diversos eventos de integração separados, extraímos as ORFs desses elementos utilizando ORFfinder (Rombel *et al.*, 2002) e alinhamos contra o gene viral identificado utilizando MAFFT v7(Katoh; Rozewicki; Yamada, 2019) e verificando se as ORFs apresentavam continuidade entre si. Esse processo foi realizado com 2 casos de cada família que houve sobreposição.

3.6 Análise de sensibilidade para EBEs

Realizamos testes baseados em dados da literatura sobre elementos bacterianos para determinar a sensibilidade da ferramenta em identificar EBEs. Utilizamos o estudo de Leclercq (Leclercq *et al.*, 2016) que descreve por técnicas de biologia molecular um evento de endogenização da bactéria *Wolbachia* WvulC no genoma do isópode *Armadillidium vulgare.* Para executar a ferramenta utilizamos a seguinte linha:

eefinder -in A_vulgare_v1.fna -mt wvulc.fa.meta -db wvulc.fa -bt armadillium_characterized_host_db.fasta -p 16 -ml genus -lm 10000 -rj 100

Foram feitos 19 testes com diferentes valores de *limit merge* (-lm) e *range junction* (-rj) a fim de chegar no resultado mais próximo ao estudo de Leclercq *et al.* O *merge level* foi colocado no nível gênero já que a bactéria possui apenas um gênero. Após as análises comparamos os resultados utilizando as posições de início e fim e o tamanho do elemento, com o descrito por Leclercq (Leclercq *et al.*, 2016).

4 RESULTADOS E DISCUSSÃO

4.1 Desenvolvimento da ferramenta

O código encontra-se finalizado e depositado no repositório do Github (https://github.com/WallauBioinfo/EEfinder), acompanhado de documentação para instalação, teste rápido e descrição dos outputs.

4.2 Diferenças na eficiência entre os modos de busca

Dentre os modos do DIAMOND tivemos o *very sensitive* como maior sensibilidade alcançada pelo algoritmo com 225 EVEs retornados em um intervalo de identidade de 16,2% a 100,0% com um tempo de 42h e 34 min. O *ultra sensitive* teve os mesmo resultados de número de EVEs e intervalo de identidade porém com um maior tempo 45h 39min. O modo *fast* foi o mais rápido com 126 regiões de EVEs, 16,9% - 100,0% de identidade e tempo total de 8h 6min (Tabela 3).

Porém o BLAST teve resultados mais sensíveis com 481 proteínas recuperadas em um intervalo de identidade de 10,2% - 100,0%, com um tempo total de 56h e 14min (Tabela 3).

Tabela 3: Resultados das comparações entre BLAST e DIAMOND

Modo	Número de EVEs	Intervalo de identidade	Tempo de execução
blastx	481	10,2 - 100,0	56h 14mim
fast	126	16,9 - 100,0	8h 6min
sensitive	221	16,2 - 100,0	38h 39min
mid sensitive	223	17,8 - 100,0	31h 25min
more sensitive	219	16,2 - 100,0	38h 17min
very sensitive	225	16,2 - 100,0	42h 34min
ultra sensitive	225	16,2 - 100,0	45h 39min

Fonte: Autor, 2024

O DIAMOND funciona utilizando um algoritmo de alfabeto reduzido e no artigo de desenvolvimento da ferramenta é relatado que não existe impacto nos resultados

em comparação com o BLAST (Buchfink; Reuter; Drost, 2021). Porém em estudos de EEs é perceptível uma grande diferença na sensibilidade, tendo uma diferença de 256 EVEs que não seriam encontrados caso utiliza-se o DIAMOND ao invés do BLAST, além do DIAMOND conseguir atingir uma identidade mínima de 16,2% em comparação com o BLAST de 10,2%. Ainda é possível utilizar o DIAMOND para triagens em genomas para determinar qual contém a maior quantidade de EEs e após tal triagem realizar a pesquisa com o melhor candidato utilizando o BLAST.

4.3 Consumo de recursos computacionais

O menor tempo de processamento do *benchmark* viral foi com 8 threads e 16 GB de RAM com uma média da triplicata de 2 horas e 20 minutos, enquanto que o maior tempo (média de 2 horas e 38 minutos) foi com 4 threads e 8 GB de memória (Tabela 4).

Tabela 4: Valores dos testes de consumo de recursos computacionais com dados virais

Número de threads	Teste 1	Teste 2	Teste 3	Média
4 threads	2 horas e 30 minutos	2 horas e 53 minutos	2 horas e 31 minutos	2 horas e 38 minutos
8 threads	2 horas e 20 minutos	2 horas e 15 minutos	2 horas e 25 minutos	2 horas e 20 minutos
16 threads	2 horas e 46 minutos	2 horas e 20 minutos	2 horas e 33 minutos	2 horas e 33 minutos
32 threads	2 horas e 25 minutos	2 horas e 39 minutos	2 horas e 40 minutos	2 horas e 34 minutos

Fonte: Autor, 2024

Em relação aos testes para EBEs tivemos o menor tempo de processamento (média de 1 minuto e 18 segundos) com 16 threads e 32 GB de memória (Tabela 5). Já o maior tempo foi em média 2 minutos e 4 segundos com 4 threads e 8 GB de memória.

Tabela 5: Valores dos testes de consumo de recursos computacionais com dados bacterianos

Número de threads	Teste 1	Teste 2	Teste 3	Média
4 threads	2 minutos e 12 segundos	2 minutos	2 minutos	2 minutos e 4 segundos
8 threads	1 minuto e 24 segundos	1 minuto e 30 segundos	1 minuto e 36 segundos	1 minuto e 30 segundos
16 threads	1 minuto e 18 segundos	1 minuto e 18 segundos	1 minuto e 18 segundos	1 minuto e 18 segundos
32 threads	1 minuto e 12 segundos	2 minutos e 18 segundos	2 minutos e 30 segundos	2 minutos

Fonte: Autor, 2024

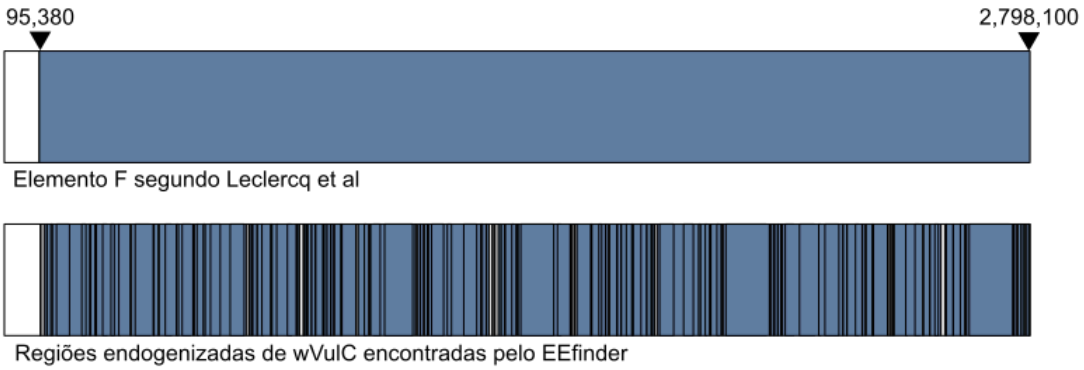
Não foi apresentado diferenças significativas no tempo de processamento pelo número de *threads* utilizado, sendo então uma ferramenta que pode ser utilizada inclusive em computadores pessoais.

4.4 Análise de sensibilidade para EBEs

Em comparação contra o estudo de Leclercq *et al*, o teste contra o genoma do isópode *Armadillidium vulgare,* utilizando as proteínas da bactéria *Wolbachia w*VulC retornou uma grande região de endogenização começando e terminando nas mesmas regiões descritas por Leclercq 95380 - 2798100 (Figura 2), esses resultados foram atingidos com os valores -lm 10000 e -rj 100.

Apesar da fragmentação apresentada no resultado do EEfinder, a ferramenta consegue entregar um resultado bastante semelhante a um método de biologia molecular, o que reflete na eficiência da ferramenta em identificar EBEs de forma reprodutível. Essa fragmentação pode ser reflexo de mutações que fragmentam os elementos em diversas regiões como acontece com EVEs, mas no resultado de Leclercq não fica claro se também foi observado esse evento visto que só descrevem o começo e final da integração, sem informações adicionais no estudo sobre a região endogenizada.

Figura 2: Representação das regiões encontradas entre os resultados de Leclercq e EEfinder, o retângulo representa um cromossomo. As regiões identificadas como *Wolbachia* *W*vulC estão em azul.



Fonte: Autor, 2024.

4.5 Análise de sensibilidade para EVEs

Os resultados que seguiram para as comparações foram com o banco de dados do Whitfield, adquiridos com o parâmetro *-ml Family*.

Os resultados utilizando as proteínas virais e o banco de proteínas do hospedeiro não caracterizadas (banco contendo proteínas *hypothetical* e *uncharacterized)* resultou em um número de EVEs de 377 (Tabela 6). Já removendo proteínas hipotéticas e proteínas não caracterizadas tivemos um número maior de elementos recuperados (578 EVEs). Quando utilizamos apenas proteínas descritas por Whitfield (Whitfield *et al.*, 2017) como banco de proteínas virais e o banco de proteínas caracterizadas o número diminui para 423 elementos encontrados, o que se assemelha ao resultado do artigo de Whitfield (472 elementos).

Tabela 6: Resultados com diferentes combinações de bancos virais e bancos de filtro.

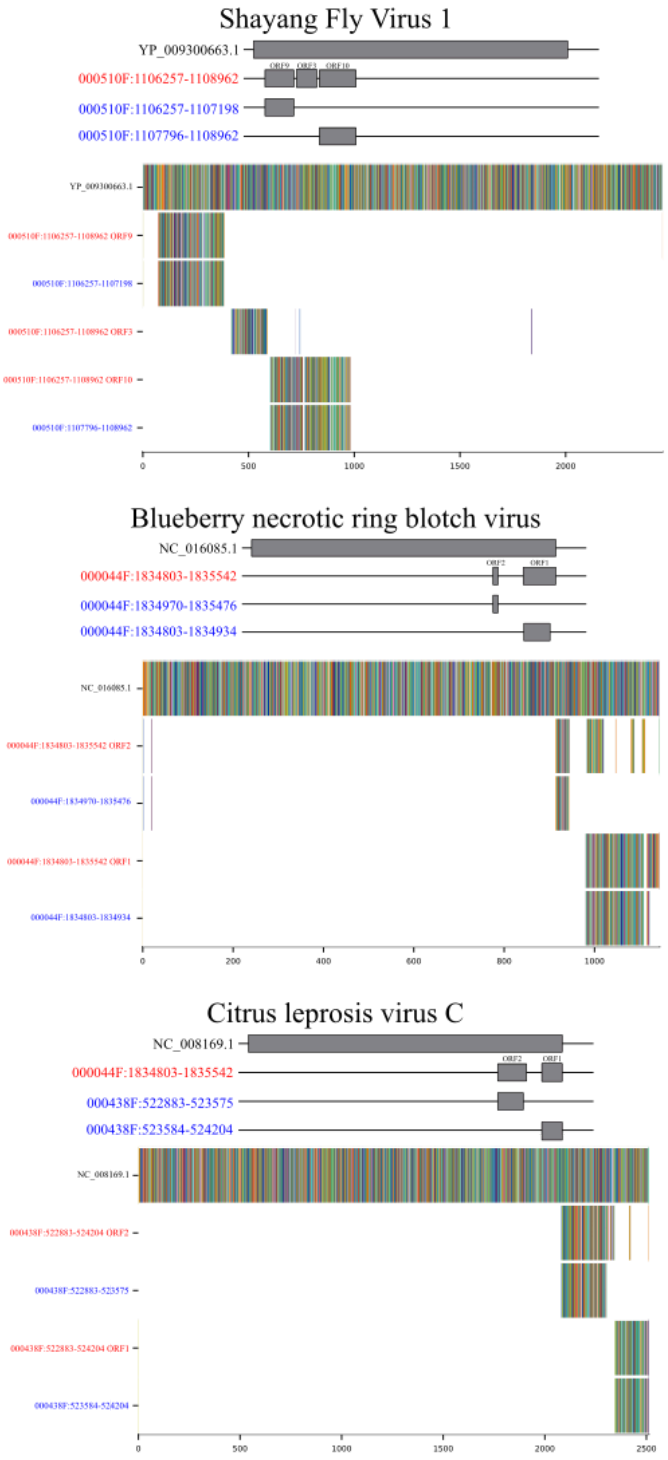
Banco viral	Banco de proteínas do hospedeiro	Número de EVEs
NCBI refseq	Banco contendo sequências anotadas como não caracterizado	377
NCBI refseq	Banco somente com proteínas caracterizadas	578
Proteínas descritas por Whitfield	Banco caracterizado	423

Fonte: Autor, 2024.

Provavelmente a diferença entre o primeiro e segundo teste com o banco viral do NCBI se deu porque diversos elementos putativos (201) foram filtrados, pois o valor de *bitscore* com as proteínas hipotéticas e não caracterizadas foi maior que o *bitscore* viral. Provavelmente essas proteínas hipotéticas e não-caracterizadas são na verdade EVEs que não passaram pela curadoria do NCBI permanecendo como não caracterizadas (uncharacterized). Esses resultados apontam a necessidade da atualização das anotações dessas proteínas, assim ajudando esse campo de pesquisa.

Para verificar se essas junções de elementos fariam sentido biológico, recuperamos as sequências e extraímos suas ORFs dos dois estudos e executamos alinhamentos múltiplos contra as proteínas que tiveram o melhor hit. Para concluir se o Whitfield realmente descreveu um elemento endogenizado degenerado como diversas endogenizações separadas. Na Figura 3 mostramos 3 exemplos desses alinhamentos com os vírus: *Blueberry necrotic ring blotch virus, Citrus Leprosis virus C* e *Shayang Fly Virus 1*. Nos 3 exemplos é possível observar que o EEfinder conseguiu juntar resultados que possivelmente quando ocorreu o evento de endogenização eram um único elemento que posteriormente sofreu fragmentação.

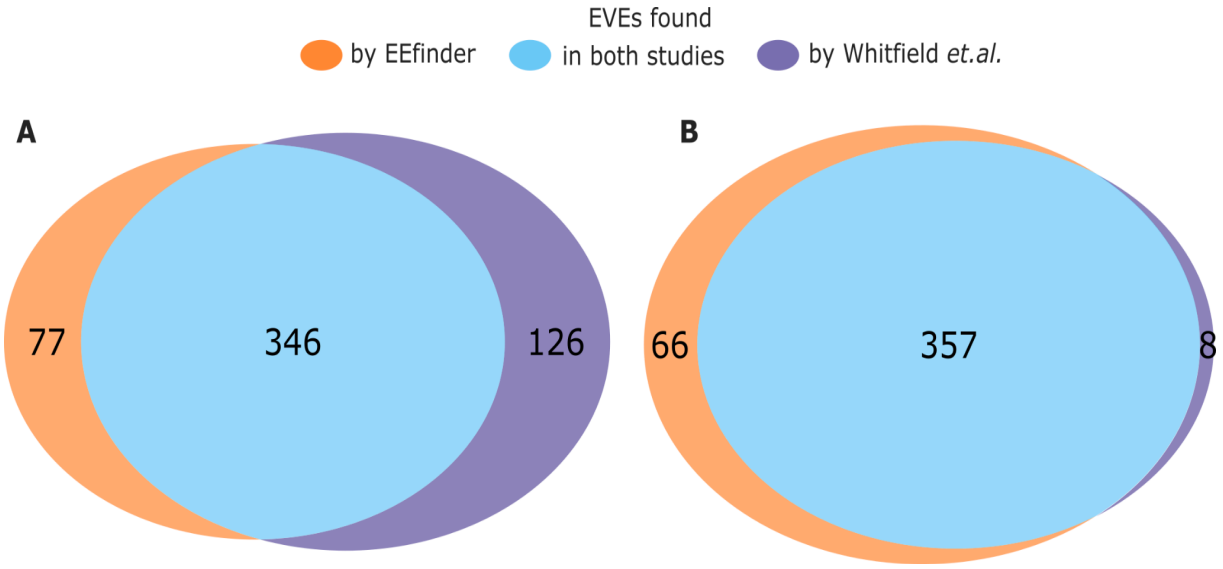
Figura 3: Alinhamento múltiplo com representação das sequências virais correspondente a região endogenizada (em fonte preta), região encontrada pelo EEfinder (em fonte vermelha), região descrita por Whitfield (em fonte azul).



Fonte: Autor, 2024.

A diferença entre o resultado do EEfinder com proteínas do Whitfield e os resultados do próprio Whitfield pode ter sido resultado da abordagem que o EEfinder emprega de juntar elementos próximos com a mesma assinatura taxonômica. Quando comparamos as regiões encontradas descobrimos que as duas análises têm 357 elementos em comum entre si, com o Whitfield apresentando 8 elementos exclusivos e o EEfinder com 66 elementos exclusivos (Figura 4B).

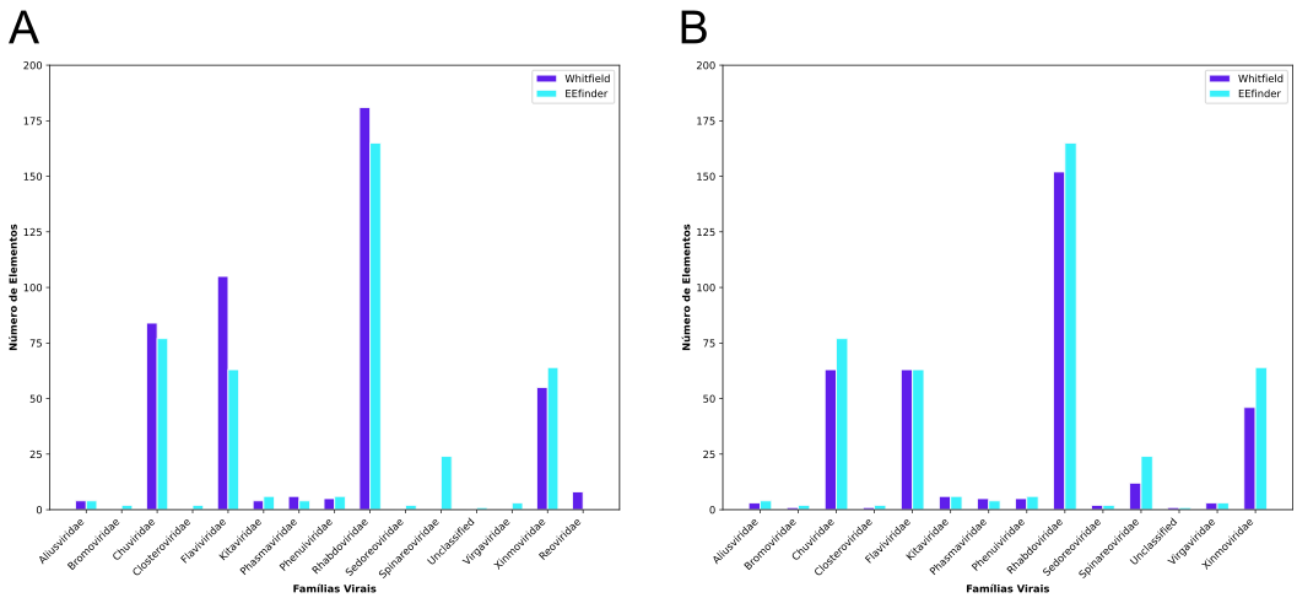
Figura 4: Gráfico de Venn com comparação das regiões endogenizadas encontradas em cada estudo. **A)** Regiões dos elementos antes da curagem de elementos degenerados **B)** Regiões dos elementos após a curagem de elementos degenerados.



Fonte: Autor, 2024.

As famílias mais presentes em ambos os estudos foram: Rhabdoviridae, Flaviviridae e Chuviridae. Das 13 famílias encontradas, 5 foram apresentadas apenas pelo EEfinder: Spinareoviridae, Virgaviridae, Sedoreoviridae, Closteroviridae Bromoviridae (Figura 5). O estudo de Whitfield apresentou uma família exclusiva (Reoviridae). O número e tamanho dos elementos encontrados pela ferramenta seguiram o mesmo padrão de distribuição apresentado pelas famílias descrito por Whitfield.

Figura 5: Distribuição dos elementos por famílias virais em cada estudo. **A)** Distribuição dos



elementos por família antes da curagem das regiões degeneradas. **B)** Distribuição dos elementos por família após curagem das regiões degeneradas.

Fonte: Autor, 2024.

O total de bases encontradas em cada estudo foi bem próximo, com uma diferença de 79,189 pb (Tabela 7). O total de bases encontradas pelo EEfinder de 405,550 pb, já Whitfield identificou 326,364 pb. Essa diferença de tamanho dos dois estudos provavelmente também é advinda da estratégia que nossa ferramenta tem de juntar elementos degenerados.

Tabela 7: Soma das bases que representam endogenizações encontradas em cada estudo.

Estudo	Bases totais encontradas que representam endogenizações
EEfinder	405,550 bp
Whitfield et. al. 2016	326,364 bp

Fonte: Autor, 2024.

Os resultados apresentados comparando ao estudo de Whitfield tem conclusões semelhantes pelo conteúdo de elementos virais endogenizados. Além de conseguir identificar elementos que foram fragmentados, o EEfinder ainda possibilita que esses resultados sejam replicados por disponibilizar arquivos *logs*. Isto demonstra a eficiência da ferramenta em identificar e classificar EVEs de forma reprodutível.

6 CONCLUSÃO

O EEfinder provou-se uma ferramenta que apresenta sensibilidade quando consegue atingir resultados tão próximos à literatura de EVEs e EBEs disponível. Além disso, o EEfinder traz consigo duas ferramentas para análise de similaridade com propósitos distintos, o DIAMOND com um tempo de execução menor e sensibilidade reduzida que pode ser utilizado para triagem de genomas e o BLASTx que apresenta uma melhor sensibilidade com objetivo de identificar todo conteúdo de EEs em um certo genoma. A ferramenta se insere no campo de EEs como uma ferramenta com baixo consumo computacional podendo ser utilizada em computadores pessoais sem um aumento do tempo de processamento. O EEfinder também traz um melhoramento metodológico da área com recursos que aumentam a reprodutibilidade do campo como: parâmetros customizáveis, arquivos de registros das análises e todas etapas automatizadas. É esperado que a ferramenta seja utilizada amplamente no estudo de elementos endógenos, podendo construir bancos de dados para refinamento dos estudos de vigilância e metagenômica, como também gerando dados para os estudos de interação patógeno-hospedeiro e evolução de patógenos.

REFERÊNCIAS

ARMEZZANI, Alessia *et al.* “Ménage à Trois”: The Evolutionary Interplay between JSRV, enJSRVs and Domestic Sheep. **Viruses**, [*s. l.*], v. 6, n. 12, p. 4926–4945, 2014.

ASAD, Hina; CARPENTER, David O. Effects of climate change on the spread of zika virus: a public health threat. **Reviews on Environmental Health**, [*s. l.*], v. 33, n. 1, p. 31–42, 2018.

ASWAD, Amr; KATZOURAKIS, Aris. Paleovirology and virally derived immunity. **Trends in Ecology & Evolution**, [*s. l.*], v. 27, n. 11, p. 627–636, 2012.

BARROWS, Nicholas J. *et al.* A Screen of FDA-Approved Drugs for Inhibitors of Zika Virus Infection. **Cell Host & Microbe**, [*s. l.*], v. 20, n. 2, p. 259–270, 2016.

BARZON, Luisa. Ongoing and emerging arbovirus threats in Europe. **Journal of Clinical Virology**, [*s. l.*], v. 107, p. 38–47, 2018.

BENEDICT, Mark Q. *et al.* Spread of The Tiger: Global Risk of Invasion by The Mosquito *Aedes albopictus*. **Vector-Borne and Zoonotic Diseases**, [*s. l.*], v. 7, n. 1, p. 76–85, 2007.

BRIEGEL, Hans; TIMMERMANN, Susanne E. *Aedes albopictus* (Diptera: Culicidae): Physiological Aspects of Development and Reproduction. **Journal of Medical Entomology**, [*s. l.*], v. 38, n. 4, p. 566–571, 2001.

BUCHFINK, Benjamin; REUTER, Klaus; DROST, Hajk-Georg. Sensitive protein alignments at tree-of-life scale using DIAMOND. **Nature Methods**, [*s. l.*], v. 18, n. 4, p. 366–368, 2021.

CALISHER, Charles H; MANESS, Kathryn S C. Development of Four Arboviruses in Mice and Application to Rapid Test Procedures. **APPL. MICROBIOL.**, [*s. l.*], v. 20, 1970.

CAMACHO, Christiam *et al.* BLAST+: architecture and applications. **BMC Bioinformatics**, [*s. l.*], v. 10, n. 1, p. 421, 2009.

CARVALHO-LEANDRO, D. *et al.* Immune transcript variations among Aedes aegypti populations with distinct susceptibility to dengue virus serotype 2. **Acta Tropica**, [*s. l.*], v. 124, n. 2, p. 113–119, 2012.

CHACON, Scott; STRAUB, Ben. **Pro git**. [*S. l.*]: Apress, 2014.

CONSOLI, Rotraut A. G. B.; OLIVEIRA, Ricardo Lourenço de. **Principais mosquitos de importância sanitária no Brasil**. [*S. l.*]: Editora Fiocruz, 1994. Disponível em: https://www.arca.fiocruz.br/handle/icict/2708. Acesso em: 16 fev. 2024.

CULICIDAE CLASSIFICATION | MOSQUITO TAXONOMIC INVENTORY. [*S. l.*], [*s. d.*]. Disponível em:

https://mosquito-taxonomic-inventory.myspecies.info/simpletaxonomy/term/6045. Acesso em: 16 fev. 2024.

DA CONCEIÇÃO, Juliana Rodrigues *et al.* Neglected tropical diseases and systemic racism especially in Brazil: from socio-economic aspects to the development of new drugs. **Acta Tropica**, [*s. l.*], v. 235, p. 106654, 2022.

DEZORDI, Filipe Zimmer *et al.* Ancient origin of Jingchuvirales derived glycoproteins integrated in arthropod genomes. **Genetics and Molecular Biology**, [*s. l.*], v. 46, n. 1, p. e20220218, 2023.

DEZORDI, Filipe Zimmer *et al.* In and Outs of Chuviridae Endogenous Viral Elements: Origin of a Potentially New Retrovirus and Signature of Ancient and Ongoing Arms Race in Mosquito Genomes. **Frontiers in Genetics**, [*s. l.*], v. 11, p. 542437, 2020.

DRAKE, John W *et al.* Rates of Spontaneous Mutation. **Genetics**, [*s. l.*], v. 148, n. 4, p. 1667–1686, 1998.

ENGELS, Dirk; ZHOU, Xiao-Nong. Neglected tropical diseases: an effective global response to local poverty-related disease priorities. **Infectious Diseases of Poverty**, [*s. l.*], v. 9, n. 1, p. 10, 2020.

FLYNN, Peter J.; MOREAU, Corrie S. Assessing the Diversity of Endogenous Viruses Throughout Ant Genomes. **Frontiers in Microbiology**, [*s. l.*], v. 10, p. 1139, 2019.

GITHUB: LET'S BUILD FROM HERE. [*S. l.*], [*s. d.*]. Disponível em: https://github.com/. Acesso em: 18 fev. 2024.

HADFIELD, James *et al.* Twenty years of West Nile virus spread and evolution in the Americas visualized by Nextstrain. **PLOS Pathogens**, [*s. l.*], v. 15, n. 10, p. e1008042, 2019.

HERNANDEZ-VALENCIA, Juan C. *et al.* A Systematic Review on the Viruses of Anopheles Mosquitoes: The Potential Importance for Public Health. **Tropical Medicine and Infectious Disease**, [*s. l.*], v. 8, n. 10, p. 459, 2023.

HOLMES, Edward C. The Evolution of Endogenous Viral Elements. **Cell Host & Microbe**, [*s. l.*], v. 10, n. 4, p. 368–377, 2011.

HUSNIK, Filip; MCCUTCHEON, John P. Functional horizontal gene transfer from bacteria to eukaryotes. **Nature Reviews Microbiology**, [*s. l.*], v. 16, n. 2, p. 67–79, 2018.

INÁCIO DA SILVA, Luísa Maria *et al.* Systematic Review of Wolbachia Symbiont Detection in Mosquitoes: An Entangled Topic about Methodological Power and True Symbiosis. **Pathogens**, [*s. l.*], v. 10, n. 1, p. 39, 2021.

KALLÁS, Esper G. *et al.* Live, Attenuated, Tetravalent Butantan–Dengue Vaccine in

Children and Adults. **New England Journal of Medicine**, [*s. l.*], v. 390, n. 5, p. 397–408, 2024.

KATOH, Kazutaka; ROZEWICKI, John; YAMADA, Kazunori D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. **Briefings in Bioinformatics**, [*s. l.*], v. 20, n. 4, p. 1160–1166, 2019.

KATZOURAKIS, Aris. Paleovirology: inferring viral evolution from host genome sequence data. **Philosophical Transactions of the Royal Society B: Biological Sciences**, [*s. l.*], v. 368, n. 1626, p. 20120493, 2013.

KATZOURAKIS, Aris; GIFFORD, Robert J. Endogenous Viral Elements in Animal Genomes. **PLoS Genetics**, [*s. l.*], v. 6, n. 11, p. e1001191, 2010.

KOUTSOVOULOS, Georgios *et al.* Palaeosymbiosis Revealed by Genomic Fossils of Wolbachia in a Strongyloidean Nematode. **PLoS Genetics**, [*s. l.*], v. 10, n. 6, p. e1004397, 2014.

LARA PINTO, Andressa Zelenski De *et al.* Novel viruses in salivary glands of mosquitoes from sylvatic Cerrado, Midwestern Brazil. **PLOS ONE**, [*s. l.*], v. 12, n. 11, p. e0187429, 2017.

LAVEN, H. Eradication of Culex pipiens fatigans through Cytoplasmic Incompatibility. **Nature**, [*s. l.*], v. 216, n. 5113, p. 383–384, 1967.

LECLERCQ, Sébastien *et al.* Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome. **Proceedings of the National Academy of Sciences**, [*s. l.*], v. 113, n. 52, p. 15036–15041, 2016.

LELAND, Diane S.; GINOCCHIO, Christine C. Role of Cell Culture for Virus Detection in the Age of Technology. **Clinical Microbiology Reviews**, [*s. l.*], v. 20, n. 1, p. 49–78, 2007.

LWANDE, Olivia Wesula *et al.* Globe-Trotting *Aedes aegypti* and *Aedes albopictus* : Risk Factors for Arbovirus Pandemics. **Vector-Borne and Zoonotic Diseases**, [*s. l.*], v. 20, n. 2, p. 71–81, 2020.

MAIA, Luis Janssen. Arbovirus surveillance in mosquitoes: Historical methods, emerging technologies, and challenges ahead. [*s. l.*],

MATTHEWS, Benjamin J. *et al.* Improved reference genome of Aedes aegypti informs arbovirus vector control. **Nature**, [*s. l.*], v. 563, n. 7732, p. 501–507, 2018.

MCCLINTOCK, Barbara. The origin and behavior of mutable loci in maize. **Proceedings of the National Academy of Sciences**, [*s. l.*], v. 36, n. 6, p. 344–355, 1950.

MI, Sha *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. **Nature**, [*s. l.*], v. 403, n. 6771, p. 785–789, 2000.

NOURI, Shahideh *et al.* Insect-specific viruses: from discovery to potential translational applications. **Current Opinion in Virology**, [*s. l.*], v. 33, p. 33–41, 2018.

PAHO/WHO DATA - ANNUAL ARBOVIRUS BULLETIN 2022 | PAHO/WHO. [*S. l.*], 2023. Disponível em: https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/annual-arbovirus-bulletin-2022.html. Acesso em: 7 abr. 2023.

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**, [*s. l.*], v. 18, n. 1, p. 512, 2017.

PALATINI, Umberto *et al.* Endogenous viral elements in mosquito genomes: current knowledge and outstanding questions. **Current Opinion in Insect Science**, [*s. l.*], v. 49, p. 22–30, 2022.

PAUPY, C. *et al.* Aedes albopictus, an arbovirus vector: From the darkness to the light. **Microbes and Infection**, [*s. l.*], v. 11, n. 14–15, p. 1177–1185, 2009.
RAMÍREZ, Ana L. *et al.* Searching for the proverbial needle in a haystack: advances in mosquito-borne arbovirus surveillance. **Parasites & Vectors**, [*s. l.*], v. 11, n. 1, p. 320, 2018.

REIS, Lúcia Aline Moura *et al.* Genus Culex Linnaeus, 1758 (Diptera: Culicidae) as an Important Potential Arbovirus Vector in Brazil: An Integrative Review. **Life**, [*s. l.*], v. 13, n. 11, p. 2179, 2023.

ROMBEL, Irene T *et al.* ORF-FINDER: a vector for high-throughput gene identification. **Gene**, [*s. l.*], v. 282, n. 1–2, p. 33–41, 2002.

RUSSO, Alice G *et al.* Novel insights into endogenous RNA viral elements in Ixodes scapularis and other arbovirus vector genomes. **Virus Evolution**, [*s. l.*], v. 5, n. 1, p. vez010, 2019.

SANJUÁN, Rafael *et al.* Viral Mutation Rates. **Journal of Virology**, [*s. l.*], v. 84, n. 19, p. 9733–9748, 2010.

SOUCY, Shannon M.; HUANG, Jinling; GOGARTEN, Johann Peter. Horizontal gene transfer: building the web of life. **Nature Reviews Genetics**, [*s. l.*], v. 16, n. 8, p. 472–482, 2015.

SOUILMI, Yassine *et al.* An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. **Current Biology**, [*s. l.*], v. 31, n. 16, p. 3504-3514.e9, 2021.

SOUZA-NETO, Jayme A.; POWELL, Jeffrey R.; BONIZZONI, Mariangela. Aedes aegypti vector competence studies: A review. **Infection, Genetics and Evolution**, [*s. l.*], v. 67, p. 191–209, 2019.

TASSETTO, Michel *et al.* Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements. **eLife**, [*s. l.*], v. 8, p. e41244,

2019.

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**, [*s. l.*], v. 93, n. 6, p. e02124-18, 2019.

TORRES-FLORES, Jesús M.; REYES-SANDOVAL, Arturo; SALAZAR, Ma Isabel. Dengue Vaccines: An Update. **BioDrugs**, [*s. l.*], v. 36, n. 3, p. 325–336, 2022.
WALLAU, Gabriel Luz. RNA virus EVEs in insect genomes. **Current Opinion in Insect Science**, [*s. l.*], v. 49, p. 42–47, 2022.

WEAVER, Scott C. *et al.* Zika, Chikungunya, and Other Emerging Vector-Borne Viral Diseases. **Annual Review of Medicine**, [*s. l.*], v. 69, n. 1, p. 395–408, 2018.
WELCOME TO PYTHON.ORG. [*S. l.*], 2024. Disponível em: https://www.python.org/. Acesso em: 14 fev. 2024.

WHITFIELD, Zachary J. *et al.* The Diversity, Structure, and Function of Heritable Adaptive Immunity Sequences in the Aedes aegypti Genome. **Current biology: CB**, [*s. l.*], v. 27, n. 22, p. 3511-3519.e7, 2017.

WILSON, Anne L. *et al.* The importance of vector control for the control and elimination of vector-borne diseases. **PLOS Neglected Tropical Diseases**, [*s. l.*], v. 14, n. 1, p. e0007831, 2020.

YE, Yixin H. *et al.* Wolbachia Reduces the Transmission Potential of Dengue-Infected Aedes aegypti. **PLOS Neglected Tropical Diseases**, [*s. l.*], v. 9, n. 6, p. e0003894, 2015.

YEN, Pei-Shi; FAILLOUX, Anna-Bella. A Review: Wolbachia-Based Population Replacement for Mosquito Control Shares Common Points with Genetically Modified Control Approaches. **Pathogens**, [*s. l.*], v. 9, n. 5, p. 404, 2020.

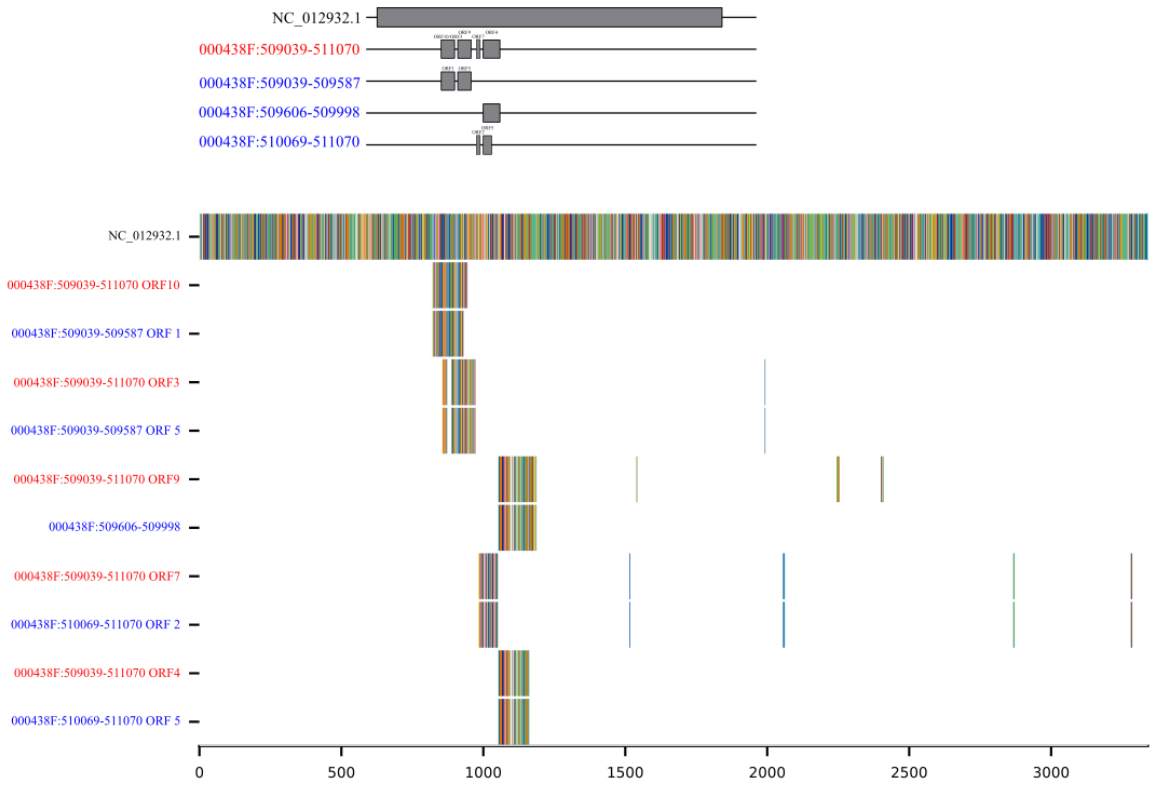
YOUNG, Paul R. Arboviruses: A Family on the Move. *In*: HILGENFELD, Rolf; VASUDEVAN, Subhash G. (org.). **Dengue and Zika: Control and Antiviral Treatment Strategies**. Singapore: Springer Singapore, 2018. (Advances in Experimental Medicine and Biology). v. 1062, p. 1–10. Disponível em: http://link.springer.com/10.1007/978-981-10-8727-1_1. Acesso em: 14 fev. 2024.

ZHANG, Yong-Zhen *et al.* Expanding the RNA Virosphere by Unbiased Metagenomics. **Annual Review of Virology**, [*s. l.*], v. 6, n. 1, p. 119–139, 2019.

ANEXOS

ANEXO A

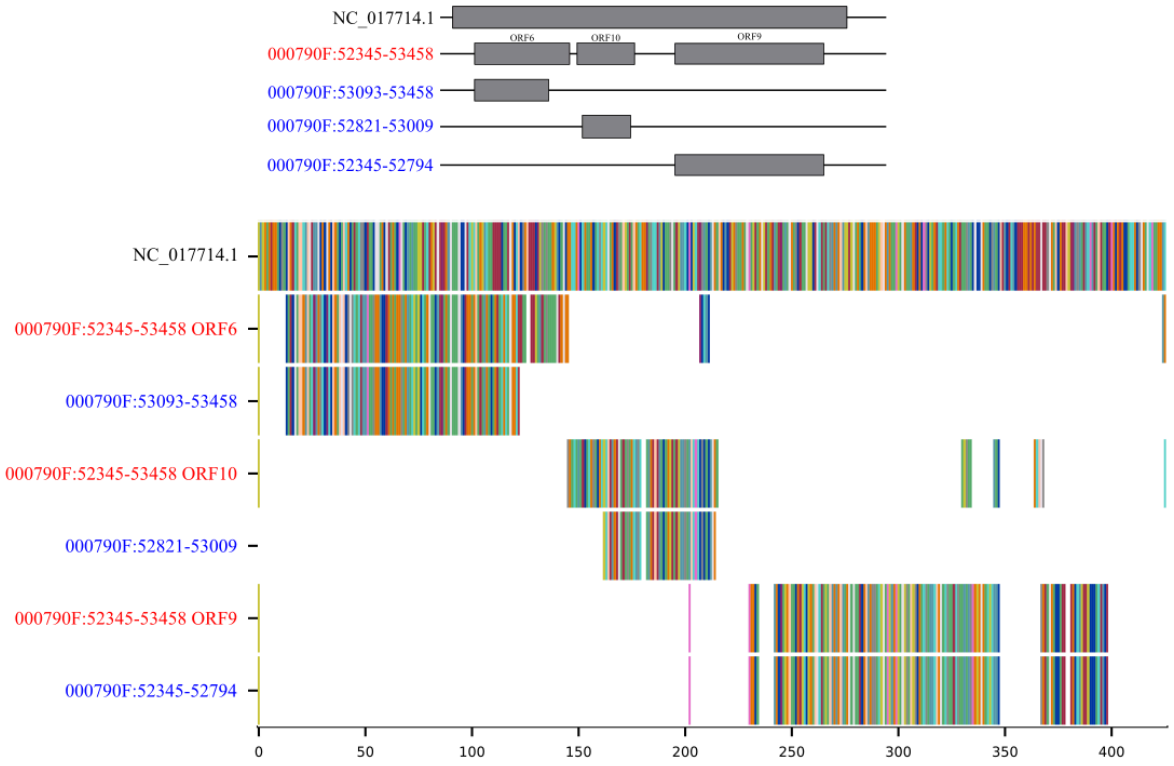
Alinhamento múltiplo da curagem com o vírus endogenizado Aedes flavivirus (em fonte preta), região encontrada pelo EEfinder (em fonte vermelha), região descrita por Whitfield (em fonte azul).



Fonte: Autor, 2024.

ANEXO B

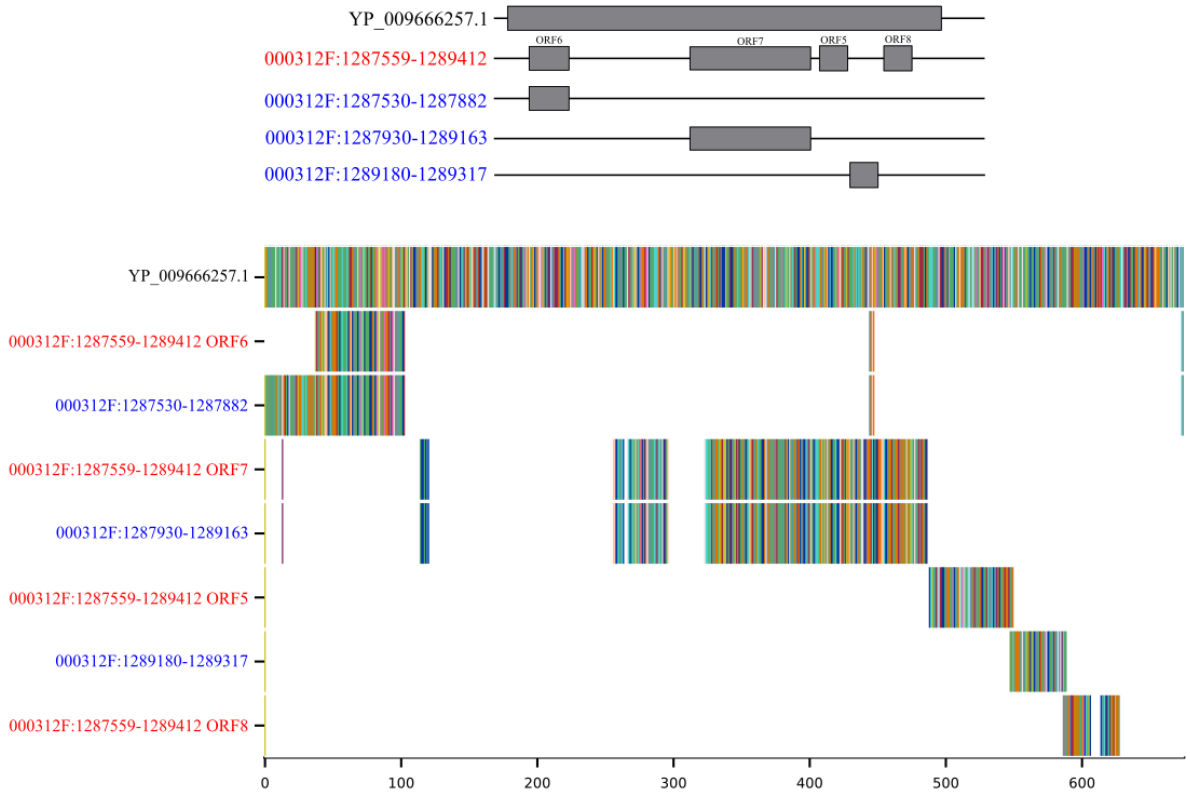
Alinhamento múltiplo da curagem com o vírus endogenizado Kotonkan virus em fonte preta), região encontrada pelo EEfinder (em fonte vermelha), região descrita por Whitfield (em fonte azul).



Fonte: Autor, 2024.

ANEXO C

Alinhamento múltiplo da curagem com o vírus endogenizado Wuchang Cockroach Virus 3 (em fonte preta), região encontrada pelo EEfinder (em fonte vermelha), região descrita por Whitfield (em fonte azul).



Fonte: Autor, 2024.