



**CENTRO UNIVERSITÁRIO DE BRASÍLIA - CEUB**  
**PROGRAMA DE INICIAÇÃO CIENTÍFICA**

**GABRIELE LOUISE TRINDADE ARAÚJO**

**ANÁLISE *IN SILICO* PARA A DETERMINAÇÃO DE CÓDIGOS DE BARRAS PARA A  
IDENTIFICAÇÃO ESPECÍFICA DE DIFERENTES ARBOVIROSES**

**BRASÍLIA**

**2021**

**GABRIELE LOUISE TRINDADE ARAÚJO**

**ANÁLISE *IN SILICO* PARA A DETERMINAÇÃO DE CÓDIGOS DE BARRAS PARA A  
IDENTIFICAÇÃO ESPECÍFICA DE DIFERENTES ARBOVIROSES**

Relatório final de pesquisa de Iniciação Científica apresentado à Assessoria de Pós-Graduação e Pesquisa.

Orientação: Paulo Roberto Martins Queiroz

**BRASÍLIA**

**2021**

## **AGRADECIMENTOS**

Honestamente, eu perdi a conta de quantas pessoas preciso agradecer, tantas foram importantes durante o processo de desenvolvimento desse trabalho que fica difícil especificar apenas algumas. Acredito que primeiramente devo agradecer à minha família como um todo por serem o motivo real de toda a minha dedicação e minha razão para ultrapassar meus próprios limites tantas vezes. Meu pai, pelo incentivo; minha mãe, pelas toneladas de apoio e admiração; meus irmãos por estarem sempre do meu lado e minha avó por ter cuidado de mim do começo ao fim. Obrigada por terem enfrentado e suportado a eu estressada de alguns dias e por terem consolado e confortado a desesperada de outros. Sem vocês eu não teria conseguido, sem vocês eu não teria nem tentado.

Quero agradecer também a cada um dos meus amigos, nesses últimos semestres eu me dei conta de quantos tenho e do quão importantes são para mim. Especialmente às minhas winx e minhas maravigolds, eu não sei o que seria de mim sem vocês. Obrigada por serem um porto seguro e por serem meu maior ponto de fuga nos dias que eu não tinha capacidade de continuar. À Drica por ter dado cada passo dessa jornada comigo, ao Edson por ter me acompanhado e me ajudado mais do que qualquer um e ao Mich por ter acreditado em mim mesmo quando nem eu acreditava.

Ao meu orientador que me mostrou o caminho que eu precisava atravessar e ao professor Raphael por me ajudar a percorrer isso da melhor forma possível. Obrigada pela paciência e por cada uma das vezes que me guiaram e inspiraram.

Por fim, agradeço a mim mesma por ter conseguido, mas saibam que foram vocês que tornaram isso possível.

## RESUMO

Arboviroses são doenças virais transmitidas por artrópodes. Há uma diversidade de infecções desse tipo disseminadas pelo mundo, dentre elas pode-se citar: Chikungunya, Dengue, Febre Amarela, Febre do Mayaro, Zika e Febre do Oeste do Nilo. Essas doenças apresentam sintomatologia muito parecida, fator que dificulta diagnósticos exclusivamente sintomáticos das infecções. Além disso, quatro dos vírus em questão comumente apresentam reação cruzada nos testes sorológicos de identificação do agente causador da infecção, afinal esses testes são pouco sensíveis e específicos. Em vista disso, testes moleculares se mostram um meio de eliminar esses problemas. Para isso, a implementação de códigos de barras, ou seja, sequências genéticas curtas e extremamente específicas, possibilita a otimização do processo de identificação de forma rápida e sensível. Objetivou-se, então, definir códigos de barras para as arboviroses citadas a fim de possibilitar melhorias em seus métodos de identificação. Através do NCBI foram coletadas sequências de genômicas completas dos vírus de interesse e dados pertinentes. Na plataforma “The Sequence Manipulation Suite” foi gerado o perfil individual de corte enzimático e esse perfil, através do programa PAST, gerou uma árvore filogenética de agrupamento. As análises estatísticas foram realizadas no programa R. Dados bibliográficos foram utilizados para selecionar genes adequados para prospecção dos códigos de barras e eles foram alinhados através do programa BioEdit. A checagem por quaisquer alterações indevidas nesse processo foi feita na ferramenta BLAST do NCBI e a obtenção dos BarCODES potenciais no programa Geneious Prime. Um novo BLAST foi realizado para avaliação de efetividade dos códigos de barras que, uma vez determinados, foram localizados através da modelagem de proteínas realizada em acesso ao SWISS-MODEL. Através da árvore filogenética foram observados perfis de aglomeração indicativos de alta estabilidade e conservação dos vírus Dengue tipo 4, Chikungunya e Zika, principalmente. Esses padrões aglomerativos indicam menor tendência a mutações, ou seja, potencialmente maior facilidade na identificação dos códigos de barras nesses grupos virais. Em vista disso, a estabilidade e conservação necessárias para seleção dos genes de prospecção foi vital para a determinação dos BarCODES. Posteriormente, nos genes selecionados para prospecção, após alinhamento e checagem por resultados inapropriados, foram determinados, pelo menos, dois códigos de barras eficientes para identificação dos vírus estudados. Os resultados menos promissores foram os de Dengue 4, Febre Amarela e Zika, ainda assim, foi possível realizar a modelagem protéica e identificação da região em que os marcadores se encontram em todos os vírus. Ainda que tenham comprimento menor que o padrão, o tamanho das regiões de identificação determinadas demonstrou ser mais apropriado para comunidades virais. Dessa forma, os elevados valores de identidade e cobertura resultantes indicaram eficiência dos marcadores definidos e podem representar maior precisão, acurácia e especificidade que os métodos convencionais de identificação. Os códigos de barras definidos aqui visam tanto proporcionar um meio de identificação mais eficiente dos vírus em questão, quanto gerar impactos positivos na epidemiologia e na taxonomia ao promover desenvolvimento do sistema de códigos de barras viral que permanece estagnado na comunidade científica.

**Palavras-chave:** Arboviroses; BarCode; Identificação viral.

## LISTA DE ABREVIÇÕES

BOLD System	BarCODE	9
CHIKV	Vírus Chikungunya	13
DENV	Vírus Dengue	13
YFV	Vírus Febre Amarela	13
MAYV	Vírus Mayaro	13
RNA	Ácido Ribonucleico	13
WNV	Vírus do Oeste do Nilo	13
ZIKV	Vírus Zika	13
Kb	Kilobase	13
DNA	Ácido Desoxirribonucleico	14
ELISA	Enzyme-Linked Immunosorbent Assay (Ensaio imunoenzimático)	14
NGS	New Generation Sequencing (Sequenciamento de Nova Geração)	15
E1	Structural Protein (Proteína Estrutural)	18
NSP2	Non-Structural Protein 2 (Proteína não estrutural 2)	18
NSP3	Non-Structural Protein 3 (Proteína não estrutural 3)	18
ORFs	Open Reading Frames (Fases de Leitura Aberta)	20
HI	Inibição da Hemaglutinação	21
NCBI	National Center for Biotechnology Information	32
HAdV7	Human adenovirus type 7 (Adenovírus Humano tipo 7)	32
CDS	Coding Sequence (Sequência Codante)	32
SMS	Sequence Manipulation Suite	33
PAST	Paleontological Statistics Software	33
UPGMA	Unweighted Pair Group Method with Arithmetic Mean (Método de grupo de pares não ponderados com média aritmética)	33
pb	Pares de bases	34
AT	Adenina – Timina	36
CG	Guanina – Citosina	37
NSP5	Non-Structural Protein 5 (Proteína não estrutural 5)	40
prM	Membrane Glycoprotein Precursor M (Precursor M da Glicoproteína de Membrana)	41

## SUMÁRIO

1 INTRODUÇÃO	7
2 FUNDAMENTAÇÃO TEÓRICA	9
2.1 AS ARBOVIROSES	9
2.2 GÊNERO <i>ALPHAVIRUS</i>	12
2.2.1 CHIKUNGUNYA	12
2.2.2 MAYARO	13
2.3 GÊNERO FLAVIVIRUS	14
2.3.1 DENGUE	15
2.3.2 FEBRE AMARELA	16
2.3.3 ZIKA	18
2.3.4 FEBRE DO NILO OCIDENTAL	19
2.4 DIAGNÓSTICOS LABORATORIAIS DAS ARBOVIROSES	20
3 MÉTODO	26
3.1 OBTENÇÃO DOS GENOMAS DAS ARBOVIROSES	26
3.2 COLETA DE DADOS BÁSICOS	26
3.3 ANÁLISES FILOGENÉTICAS DAS ESTIRPES VIRAIS	27
3.4 OBTENÇÃO DOS GENES ADEQUADOS PARA A PROSPECÇÃO	27
3.5 ALINHAMENTO DAS SEQUÊNCIAS DE BASES NITROGENADAS DO DNA	28
3.6 REALIZAÇÃO DE BLAST	28
3.7 OBTENÇÃO DOS POTENCIAIS CÓDIGOS DE BARRAS BASEADOS EM ÁCIDOS NUCLÉICOS	28
3.8 AVALIAÇÃO DA EFETIVIDADE DAS SEQUÊNCIAS CANDIDATAS COMO CÓDIGOS DE BARRAS	29
3.9 TRADUÇÃO DOS GENES PARA MODELAGEM DE PROTEÍNAS	29
4 RESULTADOS E DISCUSSÃO	29
5 CONSIDERAÇÕES FINAIS	44
REFERÊNCIAS	45
APÊNDICE A	53
APÊNDICE B	53
APÊNDICE C	73

## 1 INTRODUÇÃO

Durante décadas as arboviroses foram negligenciadas em diversos lugares do mundo por serem consideradas contribuidores menores à mortalidade global, recebendo baixa prioridade quanto aos investimentos em pesquisas e infraestrutura de saúde pública. Entretanto, as emergências e reemergências epidêmicas influenciadas pelas mudanças climáticas, globalização, urbanização e carência de manejo adequados passaram a causar real preocupação pública, tornando tanto os artrópodes transmissores das doenças quanto os vírus causadores delas, um tópico cada vez mais relevante (LIMA-CAMARA, 2016; WILDER-SMITH *et al.*, 2017).

Dentre as importantes arboviroses para a saúde pública presentes no Brasil pode-se citar: Chikungunya (CHIKV), Dengue (DENV), Febre Amarela (YFV), Febre do Mayaro (MAYV), Febre do Oeste do Nilo (WNV) e Zika (ZIKV). Os vírus causadores de todas essas infecções são compostos de um genoma de RNA fita simples sentido positivo com, aproximadamente, 11 kb de tamanho (COLPITTS *et al.*, 2012; GROCHOCKI, 2016; MEDEIROS; BRAGA, 2018; VIEIRA *et al.*, 2015; YAP *et al.*, 2017.).

Dados do Ministério da Saúde indicam que no Brasil, entre 14 de fevereiro de 2021 e 15 de maio de 2021, apenas Dengue e Chikungunya foram a causa confirmada de 109 óbitos. É necessário levar em consideração a extrema subnotificação dos casos de arboviroses ocasionadas, por exemplo, pela emergência da pandemia coronavírus, esse número potencialmente é menor que o real. No mesmo período de tempo em 2020 foram registrados 359 óbitos confirmados, dados indicam uma redução de registros de caso de, pelo menos, 57,4% (MASCARENHAS *et al.*, 2020; MINISTÉRIO DA SAÚDE, 2020; MINISTÉRIO DA SAÚDE, 2021).

Essas infecções podem apresentar quadros de gravidade significativamente amplos, mas habitualmente semelhantes, tendo em vista que grande parte das manifestações clínicas comuns se apresentam em todas elas. Destacam-se sintomas como: dores de cabeça, nos músculos e nas articulações, além de febre e erupções cutâneas, com exceção da febre amarela que geralmente apresenta amarelamento da pele e olhos ao invés da manifestação cutânea comum as outras (ESTRELA, 2017; SILVA, 2017; PAUL; VIBHUTI, 2017; PINTO-JUNIOR *et al.*, 2015; COLPITTS *et al.*, 2012).

Essa forte semelhança pode levar a complicações na identificação exclusivamente sintomática das doenças. Mesmo que elas possuam outros sintomas mais particulares que possam as diferenciar de certa forma, ainda há problemas de dúvida e confusão que impossibilitam a identificação confiável. Dessa forma, análises complementares são indispensáveis (MEDEIROS; BRAGA, 2018).

Somada a essa semelhança de sintomas, há também outro problema no desenvolvimento de diagnósticos acurados para essas patologias: a possibilidade de alguns testes sorológicos utilizados na detecção dos vírus apresentarem reação cruzada entre os anticorpos dos Flavivírus, levando a resultados de falso negativo, por exemplo. Comumente são utilizados teste ELISA ou ensaios imunocromatográficos, por serem formas acessíveis de identificação, entretanto, os testes em questão apresentam sensibilidade limitada, ainda mais em casos de infecções secundárias por dengue, por exemplo. Em conjunto a isso, os testes rápidos apresentam níveis baixos tanto de sensibilidade quanto de especificidade. No geral, os testes sorológicos só podem ser realizados dias após o início das manifestações sintomáticas de forma que sua importância se torna mais epidemiológica do que clínica, além de não serem capazes de identificar o sorotipo específico do vírus (DIAS *et al.*, 2010; KERKHOF *et al.*, 2020; LIMA-CAMARA, 2016; OLIVEIRA, 2018).

Essas complicações tornam as análises laboratoriais mais precisas ainda mais importantes, não apenas para complementar, mas para incorporar fundamentação num diagnóstico mais acurado de qual vírus em específico está causando a enfermidade. As análises moleculares de sequenciamento genético são uma ótima forma de contornar esses problemas de inespecificidade, principalmente em regiões com ocorrência simultânea das arboviroses em questão. Esses são procedimentos mais sensíveis que podem ser realizados rápida e simultaneamente com pequenas quantidades dos vírus para identificar até mesmo o sorotipo viral (GARCIA, 2018a; MEDEIROS; BRAGA, 2018; PESENTI; SILVA, 2018).

As sequências genéticas específicas do genoma viral isolado do paciente, através de análises de bioinformática, são comparadas com sequências que representam os vírus para assim identificar as espécies em análise. Uma forma de otimizar esse processo é a definição de Códigos de Barras (DNA BarCode), ou seja, uma ou algumas sequências gênicas curtas e singulares o suficiente para realizar a identificação e classificação a nível de espécie (GROCHOCKI, 2016; PURTY; CHATTERJE, 2016).

Uma das importantes premissas do método é definir fragmentos do código genético bem conservados em nível intraespecífico, a fim de permitir a segregação de espécies e até mesmo populações dentro das espécies, mas variável o suficiente para indicar divergência genética que possibilite identificações interespecíficas. Inclusive, o desenvolvimento de novas tecnologias de sequenciamento (NGS - *Next Generation Sequencing*) que prometem diminuir os custos de sequenciamento do genoma, criaram a possibilidade de utilização dos códigos de barras de DNA em novos tipos de dados genômicos, dentre eles, a rápida identificação de patógenos (BARTOLINI *et al.*, 2020; COISSAC *et al.*, 2016).

O desenvolvimento e utilização do sistema de Bar-Coding vem crescendo ao longo dos últimos anos e sequências específicas passaram a ser comumente utilizadas na identificação de diversos organismos além de serem disponibilizadas em bancos de dados *on line* como o BOLD System (BarCode of Life Data System) (<https://www.boldsystems.org/index.php>). Ainda assim, o método em questão, especificamente quando voltado para comunidades virais, permanece sendo uma vertente nova e pouco desenvolvida. Podem ser encontrados em literatura, até o momento, pouquíssimos estudos direcionados ao desenvolvimento de códigos de barras virais, ainda assim não foram desenvolvidos marcadores definitivos e nenhum deles se refere à identificação de arboviroses (CHAUDHARY; DAHAL, 2017.; PURTY; CHATTERJE, 2016).

Em vista disso, o objetivo do presente estudo foi realizar a determinação de códigos de barras específicos para identificação de seis diferentes arbovírus, de modo a facilitar e acelerar esse processo. Em conjunto a isso, foram realizadas análises de variabilidade genética intra e interespecífica com estirpes virais de diferentes partes do mundo, desenvolvendo uma análise filogenética comparativa a partir do mapeamento das sequências genômicas de interesse. Este estudo também visa identificar potenciais limitações desta ferramenta quando utilizada para identificação de vírus a fim de nortear estudos futuros e expandir o conhecimento limitado sobre o uso dessa técnica na identificação viral.

## **2 FUNDAMENTAÇÃO TEÓRICA**

### **2.1 As arboviroses**

A interação humana com outros seres durante seu processo evolutivo de expansão numérica e ocupação da Terra gerou diferentes consequências, desde prejudicar algumas

espécies ao invadir e depredar seus habitats até provocar adaptações de outras ao meio urbano criado pelo homem. Isso, por um lado, traz certas vantagens ao ser humano, como o uso do gado domesticado para a alimentação, entretanto, por outro, pode trazer problemas significativos como, por exemplo, a proliferação de insetos e as doenças que eles podem transmitir (BROWN *et al.*, 2013).

Esse é o caso de mosquitos como o *Aedes aegypti*, por exemplo, um díptero hematófago de atividade diurna originário da África, que esteve presente durante parte relevante do desenvolvimento da humanidade, se adaptando muito bem ao ambiente domiciliar, em razão de seu comportamento sinantrópico e antropofílico. A coexistência em questão acarretou despropositadamente no transporte de mosquitos para novos locais de clima favorável durante os processos de colonização e globalização do ser humano, estando por isso amplamente distribuídos no planeta (POWELL; TABACHNICK, 2013).

Graças a isso, a pressão evolutiva conjunta gerada pela interação constante tornou o ser humano um excelente meio de alimentação para esses invertebrados por ser uma fonte abundante e acessível graças a ótima adaptação que eles tiveram ao meio urbano. Isso ocorreu de forma tão significativa que estudos até mesmo indicam não apenas a preferência do mosquito por sangue humano ao invés de outros vertebrados, mas também predileção específica pelo sangue do tipo O (PRASADINI, 2019).

Essa forte presença do mosquito, somada a sua afinidade com o ser humano causam um relevante problema: a expansão das diversas arboviroses que ele pode transmitir, afinal a disseminação desses vírus está diretamente relacionada com a disseminação prévia de seu vetor (KRAEMER *et al.*, 2015).

Tendo isso em vista, o combate ao mosquito passa a ser um assunto de extrema relevância, principalmente em países de clima tropical com concentração e crescimento populacional desordenado. Esses fatores, em conjunto com as mudanças climáticas, forte globalização e falta de manejo adequado facilitam ainda mais a disseminação do vetor e, conseqüentemente, das doenças que ele transmite (LIMA-CAMARA, 2016).

Em razão disso, o controle de mosquitos como o *A. aegypti* já está sendo realizado de diferentes formas no Brasil, com inclusive diversas pesquisas a respeito de eficiência e melhorias das técnicas, desde o controle biológico até o mecânico ou químico. Inclusive, vale

ressaltar que há numerosas complicações relacionadas a esse combate, independentemente da tecnologia de controle (TERRA, 2017).

Estudos indicam que o emprego contínuo de inseticidas pode selecionar populações de mosquitos diretamente para o aumento de resistência a esses agentes, levando a uma redução significativa na eficácia deles no combate por via química, além dos diferentes efeitos degradantes que esses produtos podem causar nos sistemas biológicos (VIANA, 2020). Enquanto isso, controles físicos como destruição de criadouros impedindo a oviposição podem ser eficientes na redução populacional, mas não na eliminação já que sempre há pontos com acumulação de água que não podem ser desfeitos, por exemplo (WEERATUNGA, 2017).

Em razão disso, até que o controle eficiente do mosquito seja atingido, é necessário manter atenção quanto às doenças que eles transmitem. Dentre diversas arboviroses presentes no Brasil que podem ser citadas estão: Chikungunya (CHIKV), Dengue (DENV), Febre Amarela (YFV), Febre do Mayaro (MAYV), Febre do Oeste do Nilo (WNV) e Zika (ZIKV).

A Febre Amarela é a única dessas infecções que já possui vacina com efetividade comprovada desde 1930, as outras possuem, no máximo, vacinas ainda em estado experimental, como por exemplo a Dengvaxia® desenvolvida pela empresa Sanofi Pasteur, que apresenta 60% de eficácia e, por ser uma vacina do tipo atenuada, é contra indicada para parte relevante da população (OLIVEIRA, 2016; ROGERS, 2006). Dessa forma, considerando o manejo insuficiente e a ainda inexistência de vacinas ou medicamentos eficientes contra essas infecções, é reforçada a necessidade de aprofundamento no estudo dos arbovírus em questão. Afinal, quanto mais conhecimento a respeito deles, maior será a probabilidade de desenvolvimento de terapias efetivas de combate às enfermidades.

Um dos relevantes efeitos negativos da circulação dessas arboviroses na Saúde Pública do Brasil é o impacto econômico gerado por todo o sistema das doenças, como: os custos de diagnoses, desde os mais simples como teste rápido da Dengue até os mais complexos como PCR; custos médicos diretos, como tratamento dos enfermos; médicos indiretos, como perda de produtividade por afastamento do paciente de seu trabalho; além dos relevantes custos de combate ao vetor totalizando, em 2016, um valor de pelo menos R\$ 2,3 bilhões aos cofres públicos (TEICH, 2017).

## 2.2 Gênero Alphavírus

Dentro do gênero *Alphavirus*, que pertence à família *Togaviridae*, estão inclusos pelo menos 30 espécies de arbovírus, dos quais apenas CHIKV e MAYV são incidentes no Brasil (GROCHOCKI, 2016). Os vírus desse gênero são esféricos, com envelope viral de simetria quase icosaédrica, consistindo em uma molécula polar positiva de RNA (ácido ribonucleico) fita simples com aproximadamente 11,8 kb (YAP, 2017). Apresentam em sua composição cinco proteínas estruturais (capsídeo: C, glicoproteínas de membrana: E1 e E2, pequenos peptídeos: E3, e 6K) sintetizadas como poliproteínas e quatro não estruturais (nsP1, nsP2, nsP3 e nsP4) importantes para a replicação do vírus (LIMA-CAMARA, 2016; MOTA, 2015). Exibindo a seguinte conformação: 5' - m7G - nsP1 - nsP2 - nsP3 - nsP4 - (região de junção) - C - E3 - E2 - 6k - E1 - poly(A) - 3', sendo m7G um cap 7-metilguanossina e poly(A) uma cauda de poliadenilato (GROCHOCKI, 2016).

### 2.2.1 Chikungunya

O CHIKV é um vírus originário da África subsaariana que, em seu ciclo silvestre, circula entre seus vetores e primatas não humanos através do *Ae. albopictus* e em seu ciclo de transmissão urbana é transmitido principalmente pelo *Ae. aegypti* (HONÓRIO, 2015; WEAVER, 2015). Possui quatro diferentes genótipos relatados: Africano Ocidental, Africano Oriental Central/Sul, Asiático e o da Linhagem do Oceano Índico (ESTRELA, 2017).

Foi isolado pela primeira vez em 1952 na Tanzânia e teve seu processo de propagação iniciado com a instalação de ciclos esporádicos de transmissão, causando emergências no sudeste asiático e Índia, Quênia, Ilhas da Reunião, transmissão autóctone em países da Europa, para então atingir as Américas pelo Caribe, provocando milhares de infecções (HONÓRIO, 2015). No Brasil, os primeiros registros autóctones confirmados da doença ocorreram em 2014 (LIMA-CAMARA, 2016). Dados recentes de incidências da doença até o momento, relatados pelo Ministério da Saúde entre dezembro de 2019 e abril de 2020, apontam 80.941 casos prováveis, com 3 óbitos confirmados por critérios laboratoriais e outros 17 ainda sob investigação. No mesmo período de tempo no ano de 2021, o número de casos prováveis foi de 36.242 sem muitas investigações para constatar se houve realmente uma queda drástica no número de casos ou se foi apenas subnotificação decorrente, por exemplo, da pandemia Coronavírus.

O processo de infecção do CHIKV se dá após a picada do mosquito vetor infectado, ou seja, a realização da inoculação do vírus no corpo, rapidamente inicia-se a replicação viral nas células da pele, no local da picada, de onde os vírions se espalham para outros órgãos através da corrente sanguínea, chegando ao fígado e as articulações e, assim, caracteriza-se a fase aguda da doença (ESTRELA, 2017).

A infecção por CHIKV causa cefaléia, erupções cutâneas, mialgia e febre associada a dores articulares severas e debilitantes, tal sintoma, inclusive, foi o que deu origem a denominação Chikungunya, significando “andar curvado” em idioma africano (GROCHOCKI, 2016; ESTRELA, 2017). Essa poliartrite é característica da doença, afetando em torno de 80% dos pacientes e geralmente se apresentando nos punhos, tornozelos e cotovelos, podendo tanto melhorar em alguns dias quanto durar meses ou anos, mesmo após o término do quadro febril (DONALISIO, 2015). Além disso, esse quadro articular crônico interfere economicamente na vida do indivíduo em razão da grande perda de produtividade gerada. Ainda que seja incomum a ocorrência de casos severos da doença, manifestações neurológicas como encefalite e mielite, além de miocardite podem eventualmente levar ao agravamento dos casos ou, até mesmo, a óbitos, majoritariamente em bebês e idosos (HONÓRIO, 2015).

### **2.2.2 Mayaro**

O ciclo de transmissão do MAYV é majoritariamente enzoótico, envolvendo mosquitos transmissores e primatas não humanos nas florestas tropicais da América do Sul, de onde é provavelmente originário, acometendo desde essas regiões até países da América Central, com casos esporádicos importados em países da Europa e América do norte (FRIEDRICH-JANICKE, 2014). Seu primeiro isolamento ocorreu em Trinidad no ano de 1954, quando foi inclusive descrito, com relatos posteriores de Febre do Mayaro em vários países das Américas Central e do Sul (SUCUPIRA, 2019).

Possui dois diferentes genótipos identificados, com um deles ocorrendo exclusivamente, até o momento, no estado do Pará, sendo que, o MAYV apresenta alta taxa de mutação, aumentando potencialmente sua chance de adaptação a hospedeiros (FIGUEIREDO, 2014; SILVA, 2017).

A doença ocorre majoritariamente em áreas rurais, é diagnosticada com base apenas na sintomatologia do paciente, em razão do alto custo e falta de estrutura para realização dos

testes laboratoriais necessários nesses locais, dessa forma, é provável que o número de casos esteja sendo subestimado e a doença potencialmente negligenciada (SUCUPIRA, 2019). Outro indicativo disso são estudos realizados com amostras de sangue de mais de 300 pessoas em comunidades do Rio Negro indicando que mais de 40% dos testados possuíam anticorpos contra MAYV, demonstrando que é uma infecção comum na região, mesmo com os poucos casos reportados da doença (FIGUEIREDO, 2014).

Até o momento, há apenas um registro em literatura com diagnóstico confirmado de óbito relacionado a Febre do Mayaro, sendo esse de um paciente no México que faleceu 30 dias após a internação hospitalar apresentando hemorragia com trombocitopenia, icterícia e encefalopatia (SUCUPIRA, 2019).

A sintomatologia da Febre do Mayaro geralmente apresenta febre aguda, manchas vermelhas na pele e artralgia, extremamente comuns em outras arboviroses, o que leva a complicações no diagnóstico dessa doença negligenciada (SUCUPIRA, 2019). Pode apresentar duas diferentes fases: a aguda caracterizada por uma viremia que dura de 3 a 5 dias e a de convalescença, quando são apresentadas as artralgias e artrites que podem durar até anos, mesmo após o término do quadro febril, levando a significativa perda de produtividade do paciente (MOTA, 2015). Outros sintomas que podem ser apresentados são: mialgia, calafrios, fotofobia, cefaléia, vômito, edemas articulares, diarreias e erupções cutâneas; além de, em alguns casos, ocorrerem eventuais hemorragias nas gengivas e congestões nasais (SILVA, 2017).

### **2.3 Gênero flavivírus**

Os Flavivírus fazem parte da família Flaviridae, com pelo menos 50 espécies de arbovírus, dentre elas estão os vírus da Dengue, Febre Amarela, Febre do Nilo Ocidental e Zika (GROCHOCKI, 2016). São compostos por uma molécula de RNA em torno de 11 kb de comprimento fita simples com polaridade positiva e envelope viral, codificando uma fase de leitura aberta (ORF - *Open Reading Frame*) flanqueada por regiões 5' e 3' não traduzidas altamente estruturadas (GEBHARD, 2011; VASILAKIS, 2011). Possui simetria morfológica esférica, nucleocapsídeo icosaédrico, três proteínas estruturais (capsídeo - C, membrana - M, envelope - E) e sete proteínas não estruturais (NS1, NS2A, NS2B, NS3, NS4A, NS4B e NS5),

gerando a seguinte conformação: 5' - C - M - E - NS1 - NS2A - NS2B - NS3 - NS4A - NS4B - NS5 - 3' (BONALDO, 2017, PINHO, 2013; SAMUEL, 2016).

### **2.3.1 Dengue**

Uma das propostas de origem do vírus da dengue é a africana, vindo das florestas da África Central e Ocidental, teoria reforçada pelo fato de o primeiro registro da dengue nas Américas ser do final do século XVIII, indicando que essa chegada ao continente está possivelmente associada ao comércio de escravos da época (VASILAKIS, 2011). Além disso, há um considerável número de arbovírus semelhantes geneticamente à dengue que circulam exclusivamente na África, assim como o próprio principal vetor da doença é originário desse mesmo continente, agregando certa força a essa proposta de origem (HOLMES, 2003).

Entretanto, o fato de todos os sorotipos silvestres terem sido detectados no sudeste da Ásia, região com alta incidência de dengue, e apenas um deles ter sido detectado na África, que apresenta incidência relativa mais baixa, traz inconsistência a origem africana do vírus, demonstrando que, claramente, sua origem ainda precisa ser analisada mais profundamente (VASILAKIS, 2011). Vale ressaltar que, mesmo que relatos de surtos de doenças clinicamente semelhantes à dengue sejam feitos há séculos, apenas em 1943 e 1945 foram isolados pela primeira vez os vírus da dengue no Japão e Havaí, respectivamente (MESSINA, 2014).

O vírus apresenta o ciclo de transmissão silvestre ancestral, que ainda ocorre entre primatas não humanos e mosquitos nas florestas do sudeste da Ásia e África Ocidental; e o de transmissão urbana, em que cada um dos diferentes sorotipos da dengue, que emergiram independentemente de seus progenitores ancestrais, circulam em transmissão inter-humana contínua, graças ao estabelecimento de populações grandes o suficiente para sustentar esse ciclo (VASILAKIS, 2011).

A Dengue é uma doença de grande destaque entre as enfermidades reemergentes, sendo considerada a mais importante devido a sua significativa morbimortalidade associada (ROCHA, 2011). É uma arbovirose extremamente disseminada, estando presente em diversos países, com pelo menos 50 milhões de casos estimados anualmente e 2,5 bilhões de pessoas sob potencial risco de infecção por viverem em países onde a dengue é endêmica (NASCIMENTO, 2018).

O vírus é inoculado no ser humano através da picada do mosquito vetor infectado, iniciando sua replicação já nas células do local da picada, células musculares estriadas, lisas, fibroblastos e linfonodos, para então se disseminar pelo corpo circulando livremente no plasma sanguíneo ou no interior de algumas células do sangue, com preferência pelas fagocitárias que são seu maior sítio de replicação (OLIVEIRA, 2012; MELO, 2014).

No Brasil, a Dengue passou a ter importância epidemiológica após a primeira onda epidêmica, iniciada em 1986 no Rio de Janeiro que se estendeu pelo território do país até o Nordeste. A introdução dos quatro sorotipos presentes no Brasil geralmente estava associada a uma onda epidêmica (MEDEIROS, 2018). Dados epidemiológicos recentes, coletados entre dezembro de 2019 e abril de 2020 pelo Ministério da Saúde, indicam 979.764 casos prováveis no Brasil, sendo: 404 confirmados como casos graves, 4.785 apresentando sinais de alarme e 205 óbitos. No mesmo período do ano de 2021, esse número caiu para 348.508 casos prováveis, podendo a queda ser relacionada tanto com diminuição dos casos quanto com subnotificação dos mesmos.

Existem, detectados até o momento, 5 diferentes sorotipos da dengue (DENV 1-5) cada um possuindo entre três e cinco diferentes genótipos, sendo que os quatro primeiros sorotipos compartilham em torno de 65% de seus genomas e apresentam ocorrência simultânea no Brasil (MUSTAFA, 2015; PINHO, 2013).

A infecção pode gerar diferentes manifestações clínicas possíveis: doença indiferenciada, com sintomatologia leve ou assintomatologia; dengue clássica; febre hemorrágica; e síndrome de choque da dengue. É importante ressaltar que a infecção por qualquer um dos sorotipos promove proteção vitalícia contra ele, entretanto, aumenta a probabilidade de manifestação clínica mais grave em novas infecções por sorotipos diferentes (MONCAYO, 2004). Outros fatores potencialmente agravantes de quadros clínicos são: a cepa infectante do vírus, idade do paciente, seu histórico genético e possível influência de gênero e nutrição (VASILAKIS, 2011).

### **2.3.2 Febre Amarela**

A história de disseminação do YFV acompanha a do *A. aegypti*, ambos foram espalhados da África para diferentes países na época da escravidão, com o vírus mantendo-se

até os dias de atuais enzoótico em florestas tropicais tanto da África quanto das Américas (VASCONCELOS, 2003).

O primeiro isolamento ocorreu em 1927 na África Ocidental, mais precisamente no país de Gana e até o momento existe apenas um sorotipo reconhecido do vírus, entretanto, a variabilidade genética entre as cepas virais caracteriza sete genótipos diferentes, cinco na África e dois nas Américas, sem informações sobre qual deles é mais patogênico que outro (BONALDO, 2017).

Os primeiros registros que se tem das epidemias causada pelo VFA indicam uma onda epidêmica em 1648 no México e em Guadalupe depois, no decorrer dos séculos seguintes, epidemias regulares e extremamente alarmantes no Caribe, Américas Central e do Sul, parte dos Estados Unidos e alguns países da Europa (ROGERS, 2006). Entretanto, graças a produção de uma vacina eficaz contra o vírus em torno de 1930 e as fortes campanhas de vacinação, o ciclo urbano de transmissão, que ocorre entre mosquito vetor e seres humanos, foi drasticamente reduzido, mesmo que a transmissão ainda se mantenha, majoritariamente em seu ciclo silvestre, envolvendo mosquitos vetores e primatas não humanos (NORONHA; CAMACHO, 2017).

Vale ressaltar que, para alguns desses primatas, que são potenciais indicadores da circulação do patógeno, o VFA é altamente nocivo, causando epizootia com intervalos regulares em regiões específicas, eventualmente coincidindo com surtos esporádicos da doença em humanos não vacinados em áreas rurais (MIR, 2017). Em razão da transmissibilidade relevante, no Brasil, o combate ao mosquito precisou ser intensificado no século XX, visando conter surtos de Febre Amarela e atingindo certo sucesso no combate ao mosquito entre os anos de 1958 a 1973 (TERRA, 2017).

No Brasil, o VFA causou as primeiras relevantes ondas epidêmicas em 1685 e 1849, a partir daí foram estabelecidas as primeiras normas de campanha contra a doença no país, levando posteriormente a uma melhor organização da defesa sanitária nacional (COSTA, 2011). De acordo com o histórico epidemiológico de 1999 a 2016, é apontada uma mortalidade média de 51,8% dos casos, ressaltando a necessidade de agilidade na investigação e detecção dos casos suspeitos, assim como tratamento, acompanhamento e, principalmente, vacinação preventiva (VIEIRA, 2020).

O processo de infecção se inicia após a inoculação do vírus através da picada do mosquito infectado e pode ocorrer de diferentes formas em humanos, macacos e hamster, por exemplo. Especificamente no homem, o vírus é inoculado no hospedeiro através da picada do mosquito vetor infectado, de onde segue para os linfonodos onde inicia seu ciclo de replicação e, após a liberação dos vírions na corrente sanguínea e início do período de viremia, ele atinge o fígado (LEITE; ERRANTE, 2017).

A resposta imunológica ao VFA é ampla e variável, apresentando desde quadros clínicos oligossintomáticos até formas fulminantes (VASCONCELOS, 2003). A infecção pode causar doença viscerotrópica grave, mas, geralmente, apresenta apenas febre súbita, cefaléia, astenia e tontura; agravando-se, pode passar a expressar também náuseas, vômitos, mialgias e artralgias; e nos piores casos esses sintomas são intensificados havendo também presença de icterícia, albuminúria, oligúria e hemorragias (KLITTING, 2018).

O monitoramento da Febre Amarela no Brasil é feito anualmente, informações coletadas entre julho de 2019 e abril de 2020 indicam 2.472 mortes suspeitas de macacos, das quais: 269 foram confirmadas como epizootias por critério laboratorial, 601 foram descartadas, 378 ainda estão sob investigação e 1.224 foram classificadas como indeterminadas por impossibilidade de amostragem para diagnóstico. Já a respeito dos casos humanos, nesse mesmo período foram relatados 749 casos suspeitos, com 14 deles confirmados laboratorialmente e 3 óbitos, sendo que, nenhum dos infectados era vacinado ainda que hajam serviços de vacinação nas 27 Unidades Federativas do país, com um número estimado de 12,8 milhões pessoas não vacinadas (MINISTÉRIO DA SAÚDE, 2020).

### **2.3.3 Zika**

O ZIKV foi isolado pela primeira vez em 1947 de um macaco *Rhesus* fêmea, na Floresta Zika localizada na Uganda, originando assim o nome do vírus (HADDOW, 2012; VASCONCELOS, 2015). Estudos genéticos indicam que ele é originário da África e divergiu para três diferentes linhagens: a Africana, a Asiática e a Americana (GUBLER, 2017).

Infecções por ZIKV são reportadas desde 1960 em regiões da África e Ásia, ainda assim, a primeira importante onda epidemiológica ocorreu apenas em 2007 no Estado Federativo da Micronésia, em seguida se disseminou para algumas ilhas do pacífico entre 2013 e 2014 e

então atingiu as Américas Latina e do Norte em 2015 e 2016, respectivamente (KINDHAUSER, 2016; SONG, 2017).

O ZIKV é transmitido entre mosquitos vetores e primatas não humanos em seu ciclo silvestre e entre seres humanos e vetores em seu ciclo urbano, sendo que, ainda não há confirmações de que os primatas servem como reservatório obrigatório no ciclo urbano (GARCIA, 2018b; SONG, 2017).

A introdução do vírus no Brasil ocorreu em 2015, quando foram relatados 31.616 casos confirmados de infecção até o ano seguinte. Dados recentes da incidência da doença coletados pelo Ministério da Saúde apontam, entre dezembro de 2019 e março de 2020, 7.119 casos prováveis, ainda que nenhum dos casos tenha levado a óbitos nesse intervalo de tempo exclusivamente. Já no mesmo período em 2021 esse número foi para 2.006, sofrendo influência de subnotificação em decorrência de emergências pandêmicas.

A sintomatologia da Zika é de certa forma inespecífica e muitas infecções assintomáticas foram descritas, entretanto, são comumente apresentadas: erupções cutâneas, febres, artralgias, mialgias, astenias e conjuntivites; além de descrições de casos associados com icterícia, cefaléia, e até mesmo manifestações inespecíficas como anorexia, náuseas, vômitos, vertigem e dor retro-orbital (MUSSO, 2015; PINTO-JUNIOR, 2015). Uma maior preocupação a respeito do ZIKV está relacionada à infecção de gestantes pelo vírus, afinal, foi confirmada em 2016 a relação disso com a ocorrência de microcefalias em bebês (GARCIA, 2018b).

#### **2.3.4 Febre do Nilo Ocidental**

O vírus do oeste do Nilo foi negligenciado por décadas antes de receber a devida atenção após causar diversas emergências epidêmicas. Esse é um flavivírus de encefalite extremamente distribuído geograficamente que foi isolado pela primeira vez na África em 1937, durante vigilância de rotina voltada ao vírus da febre amarela. Especificamente na Uganda, às margens ocidentais do rio Nilo, dando origem ao nome do vírus como referência à região. Mesmo após a identificação viral, ele só foi associado às epidemias que causava no início dos anos 50 quando foi realizada sua caracterização epidemiológica (PINTO, 2015; MAY, 2011).

Já foram registrados surtos epidêmicos em diversos locais do mundo, desde a Rússia até os Estados Unidos, Caribe e Brasil. Já foi observado como a disseminação do vírus pode ocorrer não apenas rapidamente, mas também de forma preocupante como, por exemplo, quando o vírus foi introduzido na cidade de Nova Iorque em 1999 causando 21 mortes humanas antes de se espalhar para os outros estados americanos e chegando a causar, em 2010, 57 mortes e 629 manifestações neuro invasivas da doença dentre os mais de mil casos registrados (SILVA, 2016).

A infecção pelo WNV em humano é geralmente subclínica, poucos dos pacientes infectados desenvolvem sintomas da presença do vírus no organismo. Os sintomas são basicamente: quadro febril, dores de cabeça, musculares e nos olhos, além de náuseas, vômito, fadiga, manifestações cutâneas e até mesmo encefalite, mas podem se agravar em pacientes imunocomprometidos. Ainda que majoritariamente a sintomatologia não seja grave, há evidências de que a infecção por esse vírus, de forma crônica e persistente, pode causar problemas ao paciente por até sete anos (GUIMARÃES, 2021; MURRAY, 2010).

O WNV está presente no Brasil desde, pelo menos, 2011, de acordo com evidências sorológicas coletadas no Pantanal Mato-grossense. Trabalhos confirmaram a circulação do vírus em larga escala no país, ainda quando era observada uma ausência de casos clínicos. Apenas alguns anos depois, em 2014, o primeiro caso em humanos foi registrado (OMETTO, 2013; SILVA, 2016; VIEIRA, 2015).

Levando em consideração o histórico do vírus, que causou um surto de encefalite humana nos Estados Unidos, é demonstrada a necessidade de monitoramento da expansão do agente etiológico no Brasil a fim de evitar quaisquer casos semelhantes no futuro. Por isso sua identificação se torna um assunto relevante, assim como é o caso de todas as outras arboviroses aqui citadas.

#### **2.4 Diagnósticos laboratoriais das arboviroses**

A forte semelhança sintomatológica das infecções aqui estudadas pode levar à complicações na identificação exclusivamente sintomática da doença, afinal, como apresentado anteriormente, ainda que elas possuam outros sintomas particulares que possam as diferenciar de certa forma, ainda há problemas que impossibilitam a diagnose confiável. Isso agrega extrema importância à necessidade de análises laboratoriais, não

apenas para complementar, mas para conferir fundamentação para um diagnóstico mais preciso de qual vírus em específico está causando a enfermidade (MEDEIROS, 2018).

A diagnose virológica engloba diferentes métodos de detecção da infecção e a eficiência deles interfere diretamente não apenas nas questões clínicas, mas também quanto ao monitoramento e controle das doenças. Uma grande diversidade de métodos foi desenvolvida como, por exemplo, o uso de camundongos inoculados com os vírus, que para eles são letais, apresentando a desvantagem dos altos custos de mantimento de biotérios e longo tempo de isolamento. Outro método seria o isolamento em cultura celular que, além das desvantagens já citadas no método anterior, também requer a realização da técnica repetidas vezes (TERZIAN, 2008).

Considerando a complexidade desses métodos, outros tipos de diagnósticos laboratoriais se mostram mais promissores. Eles podem ser realizados de diferentes formas, seja diretamente, através de isolamento viral e testes moleculares, ou indiretamente, através de métodos sorológicos. Dentre os testes sorológicos podem ser citados ensaios de neutralização e fixação de complemento, inibição da hemaglutinação (HI), detecção viral por imunofluorescência e detecção de antígeno viral por teste ELISA, sendo HI e ELISA os métodos mais empregados e recomendados (PINHO, 2013).

HI é uma técnica rápida e de fácil execução que consiste em avaliar a capacidade de inibição de anticorpos à hemaglutinação, entretanto, exige uma qualidade de amostra colhida em fases específicas da infecção que não é fácil de obter. Além disso, as amostras sorológicas precisam ser testadas em pares com um intervalo de pelo menos 7 dias entre elas e ainda assim não apresenta capacidade de distinguir infecções de *flavivírus* (TERZIAN, 2008; SOUZA, 2019).

Já o teste ELISA realiza a detecção de Imunoglobulinas e apresenta as mesmas vantagens de ser um método fácil e rápido que, além disso, é capaz de ser utilizado num grande número de amostras. Entretanto, quando se trata dos flavivirus há o problema da reação cruzada entre os anticorpos específicos para diferentes vírus, isso pode levar a complicações no diagnóstico das doenças (LIMA-CAMARA, 2016; TERZIAN, 2008; SOUZA, 2019).

No geral, os testes sorológicos só podem ser realizados alguns dias após o início da manifestação da sintomatologia, ou seja, sua importância é mais epidemiológica do que

clínica, além de não serem capazes de identificar o sorotipo específico do vírus (DIAS, 2010). Além disso, vacinas virais como a da febre amarela também são apontadas como fatores prejudiciais na eficiência dos testes sorológicos (BORGHETTI, 2018).

Nesse momento o desenvolvimento das técnicas moleculares passa a possibilitar a eliminação dos problemas encontrados nos testes sorológicos: sensibilidade, especificidade e tempo, afinal eles detectam sequências específicas do genoma viral rapidamente (TERZIAN, 2008). É necessário pontuar que os testes moleculares não seriam um fator excludente dentro da diagnose, afinal, já foi indicado em estudo que, idealmente, deve-se associar os testes sorológicos com os moleculares para garantir uma melhor qualidade diagnóstica nas diferentes fases das infecções (AZEVEDO, 2019). Graças a essa associação entre os testes passa a ser possível a utilização correta de cada um deles nas melhores fases de diagnóstico: os testes moleculares na fase aguda para possibilitar a confirmação diferencial entre os vírus e testes sorológicos após esse período (SOUZA, 2019).

A rápida e precisa identificação do agente etiológico de patologias é de grande importância para a saúde pública. Sem isso, o tratamento dos enfermos pode demorar para se iniciar, aguardando o diagnóstico que, sem a devida acurácia, pode estar impreciso, tais fatores levam a potenciais complicações no tratamento da doença. As formas mais precisas de identificação atualmente são as análises moleculares, como, por exemplo, PCR (Reação em Cadeia da Polimerase) e suas variações, que são testes *in vitro* de replicação enzimática de regiões curtas de DNA (GARCIA, 2018a).

Esses são procedimentos sensíveis e específicos que podem ser realizados rapidamente com pequenas quantidades do vírus para identificar até mesmo o sorotipo viral. Os produtos de PCR apresentam a sequência genética de uma parte do genoma do vírus isolado do paciente que, através de análises de bioinformática, é comparada com sequências específicas de nucleotídeos que representam os vírus para assim confirmar a identificação das espécies, os códigos de barras (GROCHOCKI, 2016).

Especificamente a PCR em tempo real, uma ferramenta que vem sendo utilizada cada vez mais em laboratórios desde seu desenvolvimento, é um dos caminhos promissores para a rápida e precisa identificação. Já foi, inclusive, desenvolvido um teste multiplex de RT -qPCR chamado kit Biomol ZDC, o que demonstra a possibilidade de utilização de testes moleculares

na identificação de arboviroses mesmo que sejam considerados muito custosos atualmente (BORGHETTI, 2018).

Levando em consideração essa problemática de altos custos financeiros das técnicas de PCR, uma potencial solução para esse problema é o Sequenciamento de Nova Geração. Através das técnicas que vêm sendo desenvolvidas, através do NGS será possível sequenciar até milhões de fragmentos massiva e paralelamente, ao invés de um único, como no processo PCR seguido de eletroforese e sequenciamento. Já existem diversas técnicas que utilizam diferentes métodos para alcançar o mesmo objetivo: sequenciamento genômico rápido, extremamente preciso para identificação e que prometem diminuir os custos do sequenciamento convencional por ter um menor custo por base identificada. Já é identificado um decréscimo drástico e contínuo do custo de sequenciamento por megabase desde 2004, quando o valor que girava em torno de mil dólares caiu para menos de um, de acordo com a lei de Moore que relaciona o tempo do desenvolvimento da tecnologia com seu custo, esse foi um processo extremamente acelerado (FERREIRA, 2019; MÁLAGA, D. E. R., 2018).

Em meios de contextualização, a primeira geração do sequenciamento é o clássico método de PCR seguido de eletroforese e sequenciamento através da tecnologia SANGER. A segunda geração são outras técnicas que ainda necessitam da amplificação prévia do DNA através da PCR e a terceira geração são as plataformas desenvolvidas que dispensam a amplificação prévia, sendo capazes de sequenciar direta e individualmente o código genético. Essas novas tecnologias visam, além das vantagens já citadas, superar as limitações do sequenciamento convencional para seja possível, por exemplo, a implantação da técnica na rotina de exames laboratoriais (STROHER, 2018).

Dentre as plataformas de segunda geração que vêm sendo desenvolvidas ao longo dos anos podem ser citadas a primeira das NGS a ser lançada, uma técnica de pirosequenciamento chamada 454 e o método Illumina que tem como princípio o uso de bridge PCR para criar agrupamentos cluster que são sequenciados após seleção por marcadores fluorescentes. Já na terceira geração, se destacam as plataformas PacBio que, através da observação de marcação fluorescente, possibilita a detecção imediata das bases nitrogenadas durante a síntese do DNA e o método da MinION que é capaz de realizar sequenciamento direto através de sinais elétricos (STROHER, 2018).

Quanto mais essas técnicas se desenvolvem, mais barato e simplificado se torna o processo de sequenciamento do DNA. Entretanto, ainda que o código genético seja desvendado através dessas plataformas, ainda são necessárias sequências de comparação padrão que serão usadas para realizar a etapa da identificação. Essas regiões gênicas são chamados de Marcadores Genéticos e possuem algumas categorias: os marcadores genéticos morfológicos, que tem como premissa a capacidade de segregar características morfológicas de forma mendeliana; os marcadores genéticos e bioquímicos que utilizam, por exemplo, metabólitos secundários para realizar a distinção; os marcadores enzimáticos que foram pouco utilizados por apresentarem um baixo grau de polimorfismo; e, por fim, os marcadores moleculares que são altamente valorizados por ser capaz de explorar o polimorfismo do código genético através da identificação de distinções diretamente no DNA estudado (TURCHETTO, 2017).

Independentemente de serem marcadores híbridos, baseados em PCR ou em sequenciamento, essas regiões são extremamente importantes no momento da identificação e isso pode ser ainda mais específico quando se utiliza uma região considerada código de barras de identificação da espécie. O Código de barras de DNA ou DNA BarCode é uma técnica que foi proposta em 2003 e vem sendo desenvolvida desde então. Inicialmente a proposta era um sistema de identificação universal utilizando o gene mitocondrial Citocromo Oxidase, entretanto, nem todos os organismos possuem esse gene ou, sequer, mitocôndrias, dessa forma, esse marcador universal estaria mais ligado à espécies animais. Em vista disso, outras regiões gênicas foram demarcadas como códigos de barras para outros organismos como, por exemplo, a região ITS para alguns microorganismos ou a matK para plantas (HEBERT *et al.*, 2003; SAMPAIO, 2015; SANTOS, 2010).

Visando impactar positivamente setores como, principalmente, a taxonomia, vem sendo desenvolvida, desde 2004, uma biblioteca virtual que agrega os códigos de barras que vem sendo propostos. “The Barcode of Life Data System” é um projeto que realiza a validação dos códigos de barras no banco de dados e a disponibilização dos mesmos, além de servir como uma ponte entre comunidades científicas de todo o mundo. A designação dos barcodes seguem as premissas de derivar de uma região gênica definida, apresentar padrões de qualidade exigidos e se referir à espécies que podem ser revisadas taxonomicamente. Códigos

de barras de diversos organismos já foram propostos, validados pelo sistema e podem ser facilmente acessados (RATNASINGHAM, 2007).

Se tratando de comunidades microbiológicas o sistema de BarCODE se apresenta como algo ainda em desenvolvimento, sem marcadores universais por diversas complicações e até mesmo impossibilidades de se gerar uma sequência que consiga segregar todas as espécies. Dessa forma, há uma diversidade de regiões apontadas como marcadores para as diferentes comunidades. Como exemplos dos marcadores definidos para esses organismos podem ser citadas as regiões: ITS e COI para protozoários; RPB2 e ITS para fungos; COI e RuBisCO para algas e 16S e COI para bactérias. Quanto aos vírus, há maiores complicações (CHAUDHARY, 2017).

Sendo os seres mais abundantes na Terra, estimadamente dez vezes o número de células existentes, é de se esperar que eles apresentem uma diversidade genética extremamente ampla e complexa. Por esse motivo, a determinação de códigos de barras eficientes se torna um desafio ainda maior e poucos estudos apresentaram dados promissores (PURTY, 2016). Utilizando diferentes métodos foram determinados códigos de barras eficientes para a identificação de enterovírus humanos, vírus da gripe aviária e vírus da língua azul que afeta mamíferos de interesse econômico como gado, por exemplo (CAO, 2010; MAAN, 2011; WEI, 2011; YIN, 2011). Além disso, há também códigos de barras determinados para identificação dos vetores das doenças estudadas neste trabalho, mas não para os vírus causadores delas (SONI, 2017; ZAMORA-DELGADO, 2015).

Dessa forma, a determinação de códigos de barras específicos para os arbovírus em questão neste trabalho poderia vir a facilitar e acelerar seu processo de identificação, além de agregar análises de variabilidade genética das estirpes virais. Afinal, a partir de análises genéticas comparativas pode-se estabelecer correspondências entre genes de diferentes espécies, gerando comparações filogenéticas que possibilitam o rastreamento específico do processo evolutivo e proximidade taxonômica entre elas (HADDOW, 2012).

Sendo assim, estudos de variabilidade genética e desenvolvimento de análises filogenéticas são extremamente relevantes para o melhor conhecimento dos vírus, agregando informações essenciais até mesmo no desenvolvimento de terapias ou vacinas contra as doenças causadas pelos mesmos. Além disso, a identificação precisa e veloz de cada um deles é de grande importância para a Saúde Pública pela necessidade de um diagnóstico rápido para

que, sabendo-se qual é o agente causador da enfermidade precisamente, sejam tomados os cuidados e realizados os tratamentos adequados das pessoas acometidas pelas doenças (GROCHOCHI, 2016).

### **3 MÉTODO**

#### **3.1 Obtenção dos genomas das arboviroses**

As sequências codantes dos vírus foram obtidas através do banco de dados públicos *on line* de sequências de nucleotídeos GenBank® (<http://www.ncbi.nlm.nih.gov>) do NCBI (National Center for Biotechnology Information). As sequências de nucleotídeos dos respectivos genomas foram convertidas em formato.txt para melhor análise e acesso aos dados obtidos.

Foram selecionadas 1562 sequências gênicas codantes completas, visando maior reprodutibilidade e atendimento ao desenho estatístico das análises moleculares: 203 de DENV1, 197 de DENV2, 204 de DENV3, 199 de DENV4, 199 de CHIKV, 37 de MAYV, 196 de ZIKV, 161 de YFV e 176 de WNV. Além disso, foram coletadas outras 10 sequências de adenovírus (HAdV 7), que foi o grupo externo utilizado para realização das análises filogenéticas. As sequências utilizadas foram de vírus isolados em diversas partes do mundo para demonstrar de melhor forma a variabilidade genética dos tipos virais.

#### **3.2 Coleta de dados básicos**

Após a seleção das sequências genômicas das estirpes de cada um dos grupos de vírus (CHIKV, DENV 1-4, YFV, MAYV, ZIKV e WNV) foi realizada a coleta das informações básicas a respeito de cada um dos genomas, sendo elas: tipo viral, código de identificação da estirpe no GenBank, local de isolamento do vírus, tamanho do genoma, quantidade de CDS's e ORF's, e quantidades relativas das bases nitrogenadas na composição genômica. Todas essas informações são disponibilizadas diretamente no NCBI e podem ser acessadas no momento da coleta do genoma. Os dados foram coletados e registrados individualmente para posterior desenvolvimento de tabelas de sumarização a partir dos grupos virais para possibilitar comparações globais entre cada um dos tipos de vírus. O acesso a tabela completa está presente nos anexos.

### 3.3 Análise genética e filogenética das estirpes virais

Em acesso ao site SMS (The Sequence Manipulation Suite – disponível no endereço <http://www.bioinformatics.org/sms2/index.html>), na opção Restriction Digest, que simula a digestão das sequências inseridas por enzimas de restrição, foram inseridas as sequências de todas as estirpes virais e verificadas as enzimas que cortam os DNA's analisados em somente um ponto. Com isso foi gerado um perfil individual binário de ausência ou presença de corte único.

A partir do lançamento desses dados binários (com o número 1 representando a presença de corte único da enzima de restrição e a ausência pelo número 0) no programa PAST (Paleontological Statistics Software) (HAMMER; HARPER; RYAN, 2001) foi possível analisar a correlação e a distância entre as estirpes, utilizando-se o algoritmo UPGMA (Unweighted Pair Group Method with Arithmetic Mean) para se determinar o agrupamento entre as sequências das estirpes virais. O eixo Y da interface foi preenchido com o endereço da estirpe no GenBank e o eixo X com as enzimas de restrição que cortam o DNA de cada estirpe somente uma vez.

Além disso, o programa possibilita o encontro de árvores filogenéticas alternativas quando há congruências na matriz de entrada, associando o grau de variabilidade intraespecífica e interespecífica a partir da análise das enzimas de restrição presentes nos grupos virais. Foi utilizada especificamente a análise multivariável na opção análise de agrupamento (Cluster Analysis) com a medida de similaridade Jaccard que é um índice específico para dados binários para gerar a árvore filogenética.

Foram realizados testes estatísticos para verificar se há diferenças significativas tanto na quantidade de nucleotídeos entre os tipos virais quanto em termos de ligações de bases nitrogenadas utilizando os valores médios de ligações AT. Para a realização desses testes foi utilizado o programa R na interface R Studio (TEAM R CORE *et al*, 2013; ALLAIRE, 2012).

### 3.4 Obtenção de genes adequados para prospecção

Foi realizada uma pesquisa bibliográfica para identificação de genes de referência que são considerados conservados e estáveis em cada um dos tipos de vírus. É imprescindível que os genes avaliados sejam constantes e não altamente mutáveis ou variáveis para possibilitar

a identificação de sequências potenciais a serem definidas como códigos de barras. Foram selecionados dois genes para cada grupo viral.

### **3.5 Alinhamentos das sequências de bases nitrogenadas do DNA**

As sequências de ácidos nucleicos dos genes de interesse foram alinhadas especificamente utilizando o programa BioEdit (HALL; BIOSCIENCES; CARLSBAD, 2011). Os genes virais selecionados foram devidamente alinhados e geraram uma sequência consenso representativa de todo o grupo analisado. Foi determinado no programa uma equivalência mínima de 85% para a realização desse alinhamento. O alinhamento foi feito por etapas; primeiro, as estirpes isoladas em um mesmo país foram alinhadas e geraram a sequência consenso representativa da variabilidade genética local, esse processo foi realizado separadamente para cada um dos tipos virais. Em seguida, as sequências geradas de cada um dos países foram novamente alinhadas, gerando as sequências consenso finais de cada um dos grupos de vírus.

### **3.6 Realização de BLAST**

O BLAST (*Basic Local Alignment Search Tool*) foi realizado através do NCBI a fim de realizar a prospecção de sequências de ácidos nucleicos correspondentes às sequências consenso geradas durante o alinhamento para garantir que não houve problemas de alteração sequencial. Foi utilizada a otimização de “megablast” para seleção exclusiva de sequências altamente similares. O *cut-off* de aceitação de sequências de bases nitrogenadas foi de, no mínimo, 80% de similaridade.

### **3.7 Obtenção dos potenciais códigos de barras baseados em ácidos nucleicos**

As sequências consenso geradas durante o alinhamento foram lançadas no programa Geneious Prime (KEARSE *et al.*, 2012) da empresa Ibis Biosciences com o intuito de se efetuar o alinhamento final usando-se a plataforma ClustalW. As sequências das várias localidades foram alinhadas em conjunto para realização da identificação de possíveis regiões curtas (com uma quantidade mínima de bases de 25 pb) e conservadas de bases nitrogenadas candidatas por análise visual. A sequência consenso final individual de cada um dos grupos virais foi

registrada individualmente e a seleção das sequências potenciais foi realizada visando identificar apenas regiões com a exata sequência de ácidos nucleicos sempre que possível.

### **3.8 Avaliação de efetividade das sequências candidatas como códigos de barras**

As sequências curtas e conservadas identificadas foram então utilizadas para realização de um novo BLAST a fim de avaliar a confiabilidade das sequências como códigos de barras de DNA e assegurar que são capazes de discriminar todas as espécies e, se possível, até mesmo os sorotipos de DENV. O *cut-off* novamente foi de, no mínimo, 80% de similaridade e a identificação de eficiência foi feita através dos valores resultantes de cobertura e identidade que deveriam atingir o valor máximo de 100%.

### **3.9 Tradução dos genes para modelagem de proteínas**

Em acesso ao site SMS (The Sequence Manipulation Suite) foi realizada a tradução de todos os genes através da opção "Translate". As sequências de aminoácidos resultantes foram então registradas separadamente para a modelagem protéica. Esse processo foi desenvolvido através do site SWISS-MODEL (Disponível no endereço: <https://swissmodel.expasy.org/interactive>). Ele realiza, a partir da sequência lançada, uma prospecção por proteínas correspondentes no banco de dados e fornece, dentre outros dados, a identificação e o grau de identidade resultantes, assim como a visualização interativa da proteína modelo. Quaisquer modelagens com identificação inadequada ou identidade abaixo de 70% foram descartadas. Por fim, foi identificado em específico o local da proteína modelo em que se encontram os códigos de barras determinados utilizando o mapa protéico disponibilizado no momento da modelagem.

## **4 RESULTADOS E DISCUSSÃO**

Inicialmente, sem realizar quaisquer discriminações, foi observada a média geral da quantidade de pares de bases dos genes coletados, sendo essa média de 10.868 pb, variando entre 10.198 e 12.347 pb. As médias individuais que mais se distanciaram desse valor foram as de CHIKV e MAYV, os únicos Togavirus aqui estudados. Além disso, apenas esses dois grupos virais apresentaram mais de uma região de codificação (CDS - *coding sequence*) e fase de leitura aberta (ORFs - *Open Reading Frames*). É necessário pontuar que os vírus são compostos

de RNA, porém, o NCBI padroniza a alteração do código genético para as bases nitrogenadas encontradas em moléculas de DNA. As informações podem ser observadas com mais detalhes na tabela 1.

Tabela 1 - Dados básicos sumarizados: média da quantidade de pb de cada tipo viral, quantidade de ORF's (Open Reading Frames) e CDS's (Coding Sequences) que as estirpes apresentaram, porcentagem média da composição de ligações AT e CG dos genomas virais.

<b>Vírus</b>	<b>Média pb</b>	<b>CDS's</b>	<b>ORF's</b>	<b>% Ligação AT</b>	<b>% Ligação CG</b>
<b>DENV1</b>	10603	1	1	53,62	46,38
<b>DENV2</b>	10654	1	1	54,28	45,72
<b>DENV3</b>	10645	1	1	53,79	46,21
<b>DENV4</b>	10614	1	1	53,04	46,96
<b>CHIKV</b>	11792	2 ou 3	2, 3 ou 4	49,14	50,85
<b>MAYV</b>	11370	2 ou 3	3	48,92	51,06
<b>WNV</b>	10955	1	1	48,92	51,07
<b>YFV</b>	10916	1	1	50,00	49,98
<b>ZIKV</b>	10775	1	1	48,91	51,09

O tamanho do genoma viral exerce influência sobre diferentes características como estabilidade dos vírions e dos capsídeos, estratégias de organização e replicação, interação entre vírus e hospedeiro, variação de exposição viral à mutações e correlações filogenéticas e evolutivas (CAMPILLO-BALDERAS; LAZCANO; BECERRA, 2015; KENNEDY; PARKS, 2009).

A forma mais simples de explicar essa variação seria através da filogenia, o que também deve explicar a diferença na quantidade de ORF's e CDS's. Os Togavírus estudados, MAYV e CHIKV, são compostos por uma quantidade maior de bases nitrogenadas que os Flavivírus. É importante citar que, levando em consideração a regra de Drake que define uma relação inversamente proporcional entre tamanho do genoma e taxa de mutação por nucleotídeo, pode-se propor que esses dois vírus devem apresentar uma maior estabilidade em sua composição genômica em relação aos outros. Dessa forma, formaram-se expectativas de que sequências conservadas para esses dois grupos virais seriam encontradas mais facilmente do que para os demais (SANJUAN *et al.*, 2010).

Em relação a composição gênica, foi observada apenas em DENV uma maior prevalência de adeninas (A) e timinas (T), enquanto que os outros tipos virais apresentaram

uma pequena prevalência de citosinas (C) e guaninas (G) ou praticamente equivalência entre os tipos de ligação.

A expectativa era observar em todos os vírus uma maior prevalência de adeninas e timinas do que citosinas e guaninas, em razão de serem genomas virais dos quais se espera uma instabilidade relevante. Afinal, a prevalência de ligações do tipo AT pode levar a uma taxa de mutação mais significativa pela maior fragilidade da ligação entre essas bases nitrogenadas. Essa era uma expectativa natural, afinal genomas virais comumente são considerados muito mutáveis (ANDRADE, 2008).

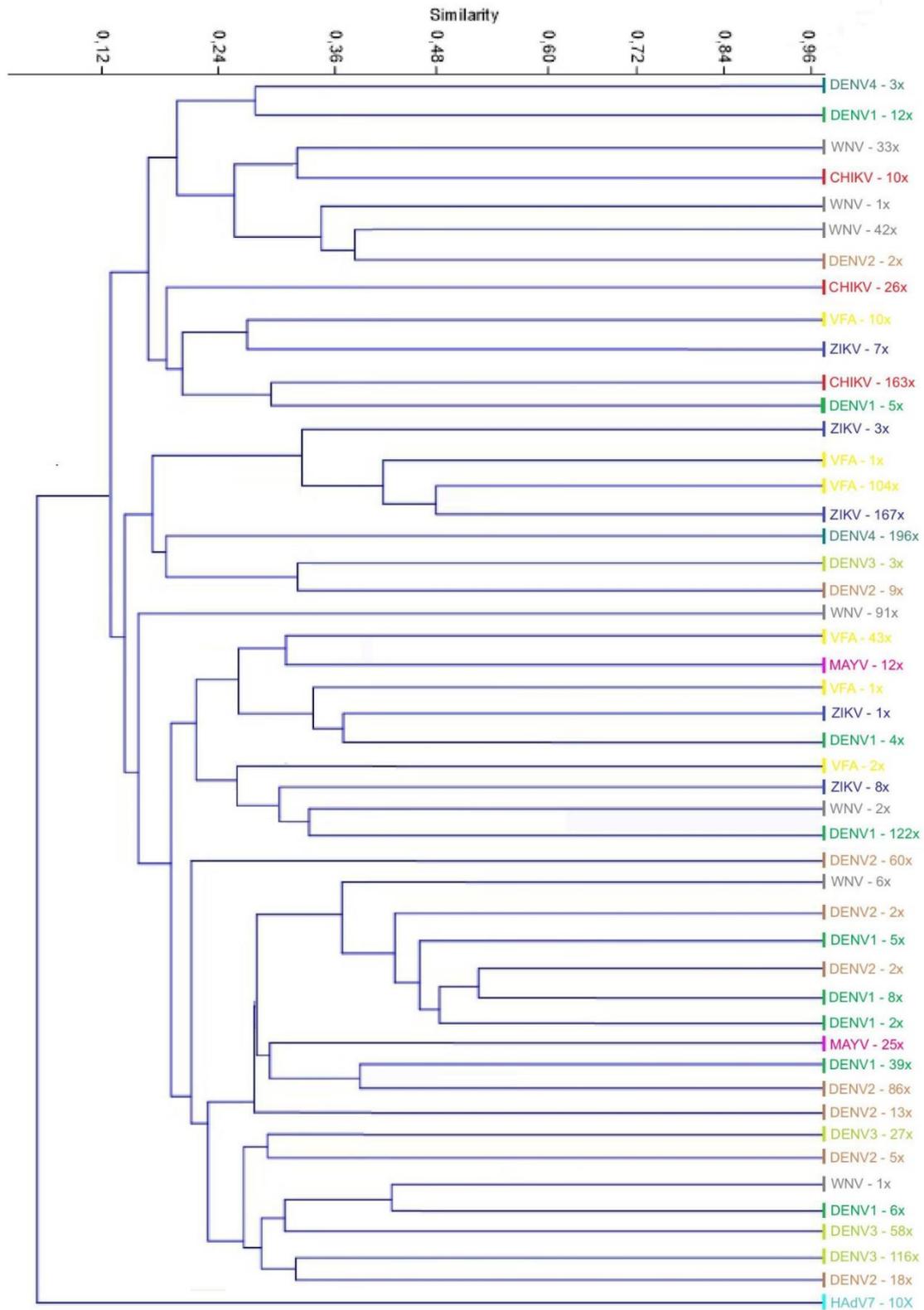
Entretanto, essa proporção só foi observada, ainda que em pequena escala, nos sorotipos do vírus da dengue, enquanto todos os outros vírus apresentaram quase equivalência entre os dois tipos de ligação ou até mesmo quase prevalência de CG. Isso poderia indicar uma tendência de mutação mais forte em DENV do que nos outros tipos virais em questão. Levando em consideração a quantidade de sorotipos que o vírus da dengue possui, pode-se apontar essa instabilidade como uma das causas de tanta variação no seu genoma. Com base nisso as expectativas iniciais eram de que haveria certa dificuldade em encontrar sequências conservadas para esses vírus. Ainda assim, não foi observada diferença estatística significativa em nenhuma dessas características, dessa forma, pode-se considerar que essa variação não é suficiente para modificar a estabilidade dos genomas em estudo.

Não foi observada nem distribuição normal dos dados (Shapiro-Wilk normality test;  $W = 0.792$ ;  $p\text{-value} < 0,05$ ), nem diferença estatística significativa entre as estirpes (Kruskal-Wallis chi-squared = 1480.2,  $df = 8$ ,  $p\text{-value} < 0,05$ ). Além disso, também não foi observada distribuição normal da quantidade de pares de bases (Shapiro-Wilk normality test;  $W = 0.78$ ;  $p\text{-value} < 0,05$ ) nem diferença estatística significativa entre os grupos (Kruskal-Wallis chi-squared = 1480.2,  $df = 8$ ,  $p\text{-value} < 0,05$ ). Com a impossibilidade de afirmar quaisquer tendências de facilidade ou dificuldade de se encontrar códigos de barras baseando-se exclusivamente nos dados analisados na tabela, foi realizada a análise através da árvore filogenética.

A partir dos dados binários do perfil de corte enzimático, utilizando as ferramentas disponibilizadas pelo programa PAST, a proximidade de cada uma das estirpes virais foi definida a fim de gerar uma árvore filogenética para representar a semelhança dos perfis. O coeficiente de correlação apresentado foi de 0,8803. Para fins de visualização, a árvore foi

comprimida e apresentada abaixo através da figura 1, mas pode ser encontrada em sua versão completa nos anexos.

Figura 1 - Árvore de proximidade gerada a partir do perfil de corte enzimático das estirpes virais.



Como esperado, a maior parte das estirpes pertencentes ao mesmo grupo viral tiveram tendência de se apresentar próximas umas das outras, indicando semelhança entre estirpes do mesmo tipo viral. Essa era uma característica já esperada, afinal, ainda que haja alta variabilidade, estirpes de uma mesma espécie ou sorotipo ainda deverão apresentar semelhanças em determinado grau. Ainda assim, foi possível observar que há variação suficiente para modificar as estirpes a ponto de seus perfis de corte enzimáticos serem mais parecidos com o perfil de outras espécies do que o de sua própria como demonstrado nos dois grupos de aglomerados de estirpes de MAYV que ficaram distantes um do outro e próximo de flavivírus.

Além disso, CHIKV e MAYV não se apresentaram em conjunto, próximos um do outro em alguma extremidade, quase como um subgrupo separado dos demais, ainda que esse fosse o esperado em razão de serem os únicos togavírus em análise. CHIKV apresentou baixa variabilidade entre suas estirpes enquanto MAYV praticamente ficou dividido em dois grupos, como se houvessem dois tipos únicos de perfil de corte que apresentaram mais proximidade com DENV e YFV do que com CHIKV. Deve-se levar em consideração que, pela pequena quantidade de estirpes analisadas em relação às outras espécies, não se pode afirmar se esse padrão continuaria presente em MAYV com uma quantidade maior de dados na análise. Além disso, existem dois genótipos (D e L) de MAYV, que podem ser os dois grupos indicados ainda que não haja registro dos dados a respeito do genótipo específico no NCBI (FIGUEIREDO; FIGUEIREDO, 2014).

O tipo viral com menor variabilidade intraespecífica foi DENV4, que manteve praticamente todas as estirpes num único aglomerado e, ainda que os outros sorotipos de dengue também tenham apresentado tendência à aglomeração, foi possível observar maior variação neles. Dessa forma, gerou-se uma expectativa de encontrar sequências bem conservadas mais facilmente em DENV4 que nos outros sorotipos. Essa indicação de maior estabilidade poderia ser reforçada pelo fato de que, dentre os sorotipos de dengue, o que apresentou menor proporção de ligações AT foi justamente o DENV4, dessa forma, numa maior escala de variação e com maiores análises essa diferença poderia vir a ser significativa.

As estirpes de ZIKV e CHIKV apresentaram um padrão semelhante ao observado em DENV4. Isso demonstra claramente como há outros fatores que influenciam na diferença de perfil de corte. Afinal, se dependesse apenas de tamanho do genoma e proporção de ligações

CG para se manter a estabilidade, todos os vírus em questão deveriam apresentar o mesmo grau de estabilidade já que a diferença entre essas características não foi significativa. Ainda assim, é claro o grau de diferenciação variado nos tipos virais.

Dessa forma, com base na variação observada através da árvore, foram criadas expectativas de encontrar maiores dificuldades no processo de prospecção de sequências principalmente para YFV, WNV e DENV1. Baseando-se, exclusivamente, no perfil aglomerativo já que foram os grupos que apresentaram maior variação. Além disso, pela impossibilidade de usar como base o tamanho do genoma e a proporção de tipos de ligação por falta de diferença estatística significativa, seriam necessários testes posteriores para identificar fatores de influência nesse quesito.

Definindo-se então expectativas de alta estabilidade de DENV4, CHIKV e ZIKV que potencialmente guiaria para uma maior possibilidade de identificação de códigos de barras eficientes, foram selecionados os genes para prospecção. Consultando a bibliografia disponível a respeito dos códigos genéticos de cada um dos tipos virais, buscou-se identificar as regiões que são consideradas mais estáveis e constantes. Ainda que os genomas virais tenham tendências mais instáveis e mutáveis que outros tipos de genoma para os quais mais facilmente se determinam códigos de barras, é inevitável considerar que eles possuem genes que apresentem essas características.

Por exemplo, genes de codificação protéica são necessários para interação com células hospedeiras, dessa forma, mutações deletérias nesses genes podem impossibilitar o ciclo viral como um todo. Ou seja, há pressão de seleção contra progênies com mutações como essa que reduzem não apenas a frequência de mutantes como também a produção de vírus num único ciclo de infecção (SANJUAN *et al.*, 2010).

Dessa forma, são nesses genes mais estáveis, teoricamente com maior presença de ligações tipo CG, que esperava-se encontrar sequências fixas que possam ser utilizadas como identificação através do processo de *barcoding*. Isso foi levado em consideração na definição dos genes selecionados como base para identificação das sequências de códigos de barras.

O gene *NSP5* é considerado um gene extremamente conservado em todos os Flavivírus, sendo essa uma região apontada como alvo potencial não apenas para identificação viral, mas também da ação de antivirais. A consulta bibliográfica a respeito de estabilidade e resistência de genes também indicou a região *NSP3* para os vírus da dengue como um

potencial marcador, assim como as regiões *NSP2* e *E1* para os Togavírus levando em consideração sua origem e correlação evolutiva para fundamentar o grau de semelhança. As mesmas características foram consideradas para todos os outros genes selecionados. Dessa forma, dentre os diversos genes que estão presentes no código genético de cada um dos vírus em análise, esperava-se que as regiões selecionadas apresentassem sequências conservadas mais facilmente (ACOSTA-AMPUDIA *et al.*, 2018; DUAN *et al.*, 2017; DUBANKOVA; BOURA, 2019; FUMAGALLI *et al.*, 2019; OLIVEIRA *et al.*, 2014; SOLIGNAT *et al.*, 2009; SCHUFFENECKER *et al.*, 2006; YOSHII *et al.*, 2012). Os genes selecionados foram sumarizados na tabela abaixo.

Tabela 2 - Regiões selecionadas para alinhamento e definição de sequência de potenciais BarCodes.

<b>Gene</b>	<b>E1</b>	<b>NSP2</b>	<b>NSP3</b>	<b>NSP5</b>	<b>prM</b>
<b>Vírus</b>	CHIKV	CHIKV	DENV1-4	DENV1-4	YFV
	MAYV	MAYV	WNV	YFV	ZIKV
				ZIKV	
				WNV	

Após o alinhamento das sequências dos respectivos genes de cada uma das estirpes dos diferentes tipos virais foi necessário garantir que não ocorreram quaisquer erros ou modificações indevidas no processo. Para isso, a realização do BLAST das sequências consenso geradas no alinhamento demonstrou variação de cobertura entre 99% e 100%, e de identidade entre 84% e 100%. Ainda assim, a maior parte dos valores se manteve em 100% tanto de cobertura quanto de identidade.

Dessa forma confirmou-se que as sequências continuaram dentro do padrão esperado para cada um dos tipos virais, isentos de erros ou modificações indevidas. A tabela abaixo demonstra a variação dos valores de identidade e cobertura resultantes do BLAST de cada um dos tipos virais:

Tabela 3 - Variação das porcentagens de cobertura e identidade resultante do BLAST das sequências consenso.

<b>Vírus</b>	<b>Região</b>	<b>Cobertura</b>	<b>Identidade</b>
CHIKV	E1	100%	98 - 100%
	NSP2	99 - 100%	99 - 100%
DENV1	NSP3	99 - 100%	95 - 100%

	NSP5	99 - 100%	96 - 100%
DENV2	NSP3	99 - 100%	94 - 100%
	NSP5	100%	94 - 100%
DENV3	NSP3	100%	98 - 100%
	NSP5	100%	98 - 100%
DENV4	NSP3	99 - 100%	96 - 100%
	NSP5	100%	87 - 100%
MAYV	E1	100%	98 - 100%
	NSP2	100%	97 - 100%
WNV	NSP3	99 - 100%	84 - 100%
	NSP5	99 - 100%	84 - 100%
YFV	NSP5	99 - 100%	85 - 100%
	prM	99 - 100%	93 - 100%
ZIKV	NSP5	99 - 100%	95 - 100%
	prM	100%	99 - 100%

Faz-se necessário apontar a alta taxa de cobertura e identidade resultante do BLAST de prospecção das sequências-consenso geradas pelo alinhamento desses genes. Ainda que tenha sido observada variação, a maior parte dos grupos virais e os respectivos genes analisados sequer se aproximaram do valor mínimo (80%) definido para aceitação das sequências e todos mantiveram correspondência com seu tipo viral. Só foi possível observar valores abaixo de 90% em DENV4, WNV e YFV. Demonstrando que os consensos gerados pelo alinhamento chegaram a se diferenciar um pouco das sequências na base de dados de realização do BLAST, mas não o suficiente para desvalidar nenhuma delas.

É preciso levar em consideração a ideia de que há outros fatores influenciando a baixa de valor nesses tipos virais. Afinal, em DENV4, por exemplo, esperava-se uma estabilidade elevada com base em seu padrão filogenético apresentado pelo perfil de corte enzimático. Ainda assim, a variabilidade das sequências utilizadas gerou consensos com valores de identidade menores que o esperado para esse tipo viral.

Através do alinhamento dessas sequências-consenso geradas foi iniciada a prospecção dos potenciais códigos de barras. A variação de comprimento dos genes selecionados para análise foi de 493 a 2791 pb. Ao levar em consideração a amplitude dimensional que os genes

virais podem ter, de menos de mil pares de bases até mais de milhões, é possível considerar que o genoma de seus hospedeiros é ainda maior, dessa forma, relativamente, os genes virais são menores que de outros organismos (CAMPILLO-BALDERAS; LAZCANO; BECERRA, 2015).

O primeiro código de barras a ser proposto foi um fragmento de 648 bp do gene mitocondrial (Citocromo Oxidase 1) para identificação de animais. A partir dessa proposta, outras regiões gênicas foram apontadas como marcadores de identificação para diferentes organismos. Esse tamanho de fragmento já é maior que o gene *prM*, por exemplo, que sequer chegou a 600 pb. Tendo isso em vista, o tamanho que os códigos de barras de identificação normalmente apresentam parece ser inviável quando aplicado à comunidade viral (COISSAC *et al.*, 2016; HEBERT *et al.*, 2003).

Os mini-barcodes tornam-se então mais apropriados para a identificação viral, afinal, as sequências podem ter de menos de 100 pb até pouco mais de 300 pb. Com base nisso, a prospecção por regiões conservadas em todas as sequências consenso gerou uma variedade considerável de sequências candidatas, ultrapassando as expectativas. Esperava-se que, com base nos fatores citados como variabilidade genética, exposição à mutação e instabilidade gênica, encontrar fragmentos altamente conservados em todas as sequências fosse ser um grande desafio (HAJIBABAEI; MCKENNA, 2012).

Foram identificadas, no total, noventa e duas sequências candidatas. Ainda assim, é necessário pontuar que a maior parte delas foi desconsiderada no momento de validação através da avaliação da eficiência de identificação das mesmas pelo do BLAST final. Os critérios mais rígidos do BLAST final eliminaram quaisquer sequências que não apresentassem 100% tanto de cobertura quanto de identidade além, obviamente, da correspondência a seu grupo viral. Essa eliminação massiva de sequências seguiu as expectativas, tendo em vista os desafios de definição de códigos de barras para vírus em razão das características genéticas apontadas anteriormente. As sequências não eliminadas, ou seja, as que apresentaram capacidade de identificar com sucesso e precisamente seus respectivos tipos virais são apresentadas na tabela 4.

Tabela 4 - Sequências de identificação finais de cada um dos tipos (códigos de barras).

Vírus	Gene	BarCODE 1	BarCODE 2
CHIKV	E1	TACGAACACGTAACAGTGATCCCGAACAC GGTGGGAGTACCGTATAAGACTCT	AAGCTCCGCGTCCTTTACCAAGGAAA

	NSP2	CAGTGCGGCTTCTTCAATATGATGCAGATG AAAGT	TCGTTGCATTACGAAGGCAAAATGCG CAC
DENV	NSP3	CTTCCAGCCATAGTCCGTGAGGCCATAAAA AG	GGGGAAGGAACAACCAGGTGTTGG AGGAGAACATGGACGTGGAGAT
1	NSP5	TGGGTTTCATGTGGAACAGGAAACATTGT GTC	AAAACATGGGCCTATCATGGATCATAT GAGGT
DENV	NSP3	CCATTCCTCAGAGCAATGCACCAAT	GCTGAGAGGGTTATAGACCCCAGACG
2	NSP5	GGGAACATAGTGTCATCAGTGAACATGAT TTCAAG	GCTGTGTCGCGAGGCTCAGCAAACACT GAGATGGTTTCGT
DENV	NSP3	CGTGAAAAAAGATCTGATTTTCATACGGAG GAGGATGGA	CAAAGGAGAGGGAGAGTTGGCAGGA ACCCACAAAA
3	NSP5	GCCACAGGCTCAGCCTCCTCCATGATAAAT GGAGTCGTGAAACT	ATGATGGGCAAGAGAGAGAAGAAAC TTGGAGAGTTTGGCAAAGC
DENV	NSP3	CCCCAGAGCAACAGCCCAATAGAAGA	X
4	NSP5	CCATACAGAACCTGGGCGTATCATGG	X
MAYV	E1	AATCAGACTGTCGAGGCGTATGTGAACGG TGACCATGCCGTAACGATTGCCGGGACGA AGTT	TCGACCGCATCAGCCGCCCTTCCTTC ATAGTATCCGTCTGCAGCTC
	NSP2	CACGAAGTCATGACAGCCGCTGCATCCCA GGGG	GGAGATCCCTGGATCAAACTTTGAG CAACAT
WNV	NSP3	TGGAAATGAGATCGTTGATGTCATGTGCC ATGCCAC	GTTTGGTTTGTTCGAAGTGTGAAAATG GGAAATGAGA
	NSP5	TCTCCAGGGGCTGGATGGAATGTGCGCGA CAC	TGGATTGAGGAGAATGAGTGGATGG AAGACAAAAC
YFV	NSP5	CTTGTGTGTACAACATGATGGGGAA	ATGGAAATGACATACAAGAACAAAGT
	prM	X	X
ZIKV	NSP5	ACCTGGTCAATCCATGGAAAGGGAGAATG GATGAC	AGAGTGTGGATTGAGGAGAACGACC A
	prM	X	X

Legenda: X indica que não foi possível estabelecer o BarCODE.

Foi demonstrado que, diferentemente do que a bibliografia indica, o gene *prM* não apresentou estabilidade e conservação suficientes para gerar qualquer sequência de identificação. Dentre os fatores que diferenciam esse gene especificamente dos demais, o comprimento é o que pode mais facilmente ser apontado, afinal, é o único com aproximadamente 500 pb, todos os demais são compostos por, no mínimo, 2000 pb. Dessa

forma, é possível supor que, ainda que a quantidade de pares de bases não tenha exercido influência estatisticamente significativa no genoma completo, talvez avaliando-se o tamanho dos genes essa diferença seja mais relevante. Ainda assim, para afirmar essa influência seriam necessárias análises posteriores.

De qualquer forma, foi possível a aprovação de uma pequena parcela das sequências de identificação candidatas iniciais, o suficiente para demonstrar a viabilidade de identificação de códigos de barras eficientes para comunidades virais. Os códigos de barras definidos foram capazes de apresentar alta capacidade de identificação e, tendo em vista o constante desenvolvimento de técnicas moleculares e seus diferentes usos, podem ser utilizados, também, para diferentes propósitos, desde identificação e taxonomia até uso no estudo e desenvolvimento de antivirais e terapias gênicas, por exemplo.

Uma vez identificados os códigos de barras, objetivou-se definir em que local ele está localizado dentro da proteína codificada pelo gene onde a região de identificação se encontra. Para isso foi realizada a modelagem proteica dos genes aqui estudados e posterior localização da região de identificação. As proteínas codificadas geradas durante o processo de modelagem podem ser visualizadas na Tabela 5.

Tabela 5 - Gene utilizados na modelagem de proteínas, grupos funcionais resultantes da prospecção e identidade perspectiva das sequências gênicas com a proteína modelo.

<b>Vírus</b>	<b>Gene</b>	<b>Identificação</b>	<b>Identidade</b>
CHIKV	E1	E1 glycoprotein Electron Cryo-Microscopy Of Chikungunya VLP	96,13 - 100 %
CHIKV	NSP2	Protease nsP2 Crystal structure of nsP2 protease from Chikungunya virus in P212121 space group at 2.59 Å (4molecules/ASU).	97,20 - 100 %
DENV1	NSP3	Serine Protease/NTPase/Helicase NS3 A Second Confirmation of the NS3 Protease-Helicase from Dengue Virus	76,05 - 76,54 %
DENV1	NSP5	RNA-directed RNA polymerase NS5 Crystal structure of full-length NS5 from dengue virus type 3	81,41 - 81,97 %
DENV2	NSP3	Serine Protease/NTPase/Helicase NS3 A Second Confirmation of the NS3 Protease-Helicase from Dengue Virus	76,54 - 77,35 %
DENV2	NSP3	Peptidase S7 Crystal structure of full length NS3	76,86 - 77,67 %

		protein (eD4NS2BNS3) in complex with Bovine Pancreatic Trypsin Inhibitor	
DENV2	NSP5	NonStructural Protein 5 Dengue Virus Nonstructural Protein 5	96,33 - 99,67 %
DENV3	NSP3	Peptidase S7 Crystal structure of full length NS3 protein (eD4NS2BNS3) in complex with Bovine Pancreatic Trypsin Inhibitor	79,38 - 80,03 %
DENV3	NSP3	Serine Protease/NTPase/Helicase NS3 A Second Confirmation of the NS3 Protease-Helicase from Dengue Virus	78,90 - 79,55 %
DENV3	NSP5	RNA-directed RNA polymerase NS5 Crystal structure of full-length NS5 from dengue virus type 3	98,54 - 99,10 %
DENV3	NSP5	Genome polyprotein Dengue 3 NS5 protein with compound 27	98,21 - 99,55 %
DENV4	NSP3	Peptidase S7 Crystal structure of full length NS3 protein (eD4NS2BNS3) in complex with Bovine Pancreatic Trypsin Inhibitor	97,57 - 99,84 %
DENV4	NSP3	Serine Protease/NTPase/Helicase NS3 A Second Confirmation of the NS3 Protease-Helicase from Dengue Virus	97,25 - 99,19 %
DENV4	NSP5	Genome polyprotein Dengue 3 NS5 protein with compound 27	76,52 - 77,30 %
DENV4	NSP5	NonStructural Protein 5 Dengue Virus Nonstructural Protein 5	74,39 - 75,61 %
MAYV	E1	Structural polyprotein Crystal structure of the homotrimer of fusion glycoprotein E1 from Semliki Forest Virus.	75,06 - 75,06 %
MAYV	NSP2	Helicase Viral helicase protein	76,03 - 76,67 %
WNV	NSP3	Flavivirian Protease NS2B Regulatory SubUnit, Flavivirin Protease NS3 Catalytic SubUnit Crystal Structure of the NS3 Protease-Helicase from Murray Valley Encephalitis Virus	79,25 - 80,39 %
WNV	NSP5	Polyprotein Crystal Structure of the full-length Japanese encephalitis virus NS5	80,75 - 81,42 %
YFV	NSP5	Genome polyprotein Crystal structure of Yellow fever virus polymerase NS5A	94,80 - 99,89 %

ZIKV	NSP5	Genome polyprotein NS5 protein	Crystal structure of Zika Virus	95,55 - 100 %
------	------	-----------------------------------	---------------------------------	---------------

É necessário levar em consideração que o banco de dados utilizado para a realização da modelagem de proteínas é limitado e pode não possuir a proteína do vírus correto. Em vista disso, faz-se necessário utilizar o modelo mais semelhante possível como o vírus codificado de um outro vírus pertencente à mesma família ou gênero.

Dentro dos grupos funcionais, ou seja, das proteínas codificadas, se encontra a região codificada pelos códigos de barras previamente definidos, totalizando 30 barcodes. A localização dessas regiões foi realizada com o objetivo de possibilitar uma observação visual dos resultados obtidos. Foram destacadas especificamente as regiões onde os códigos de barras se encontram em suas respectivas proteínas. Serão apresentados nas figuras abaixo um barcode representante de cada grupo viral, mas todos estão disponíveis nos anexos.

Figura 2: CHIKV - E1 - BarCODE 1

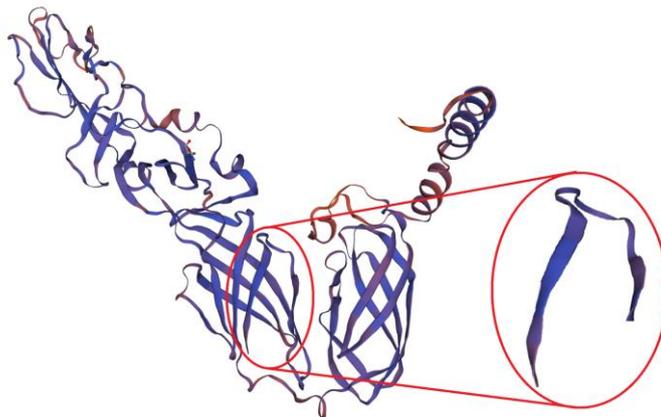


Figura 3: CHIKV - NSP3 - BarCODE 1

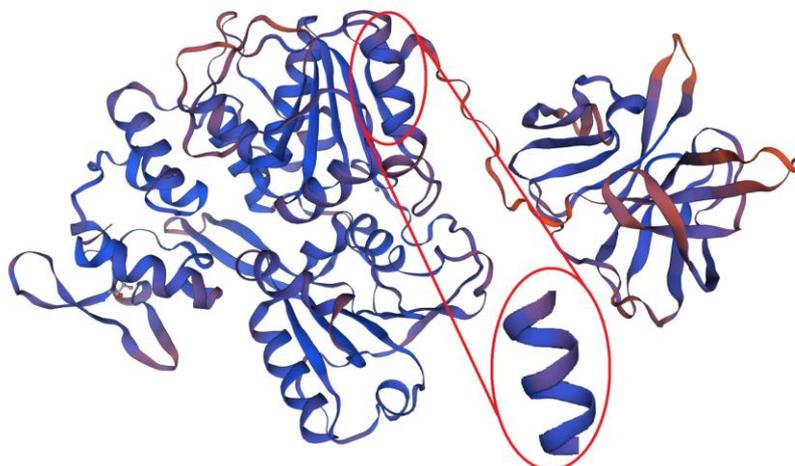


Figura 4: DENV2 - NSP5 - BarCODE 2

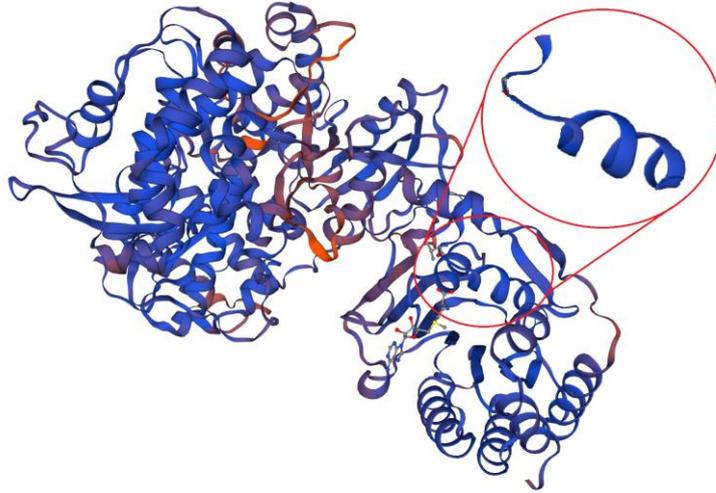


Figura 5: DENV4 - NSP3 - BarCODE 1

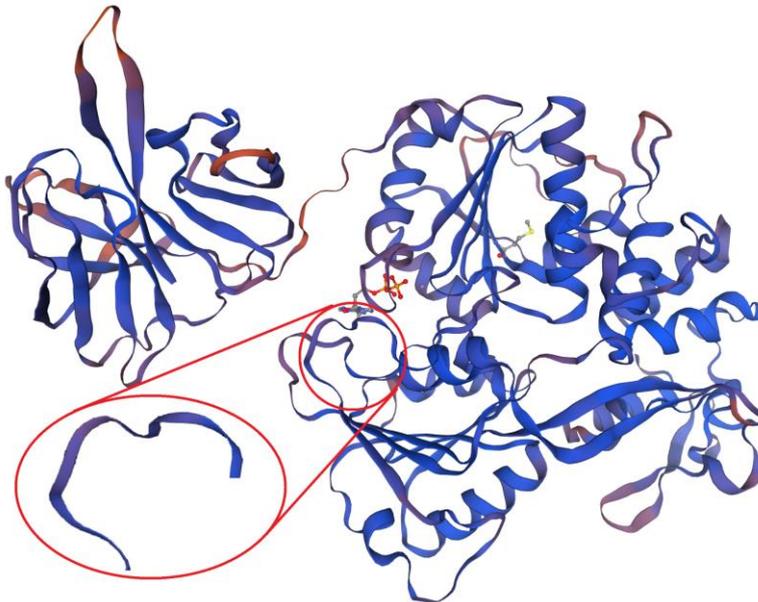


Figura 6: MAYV - NSP2 - BarCODE 1

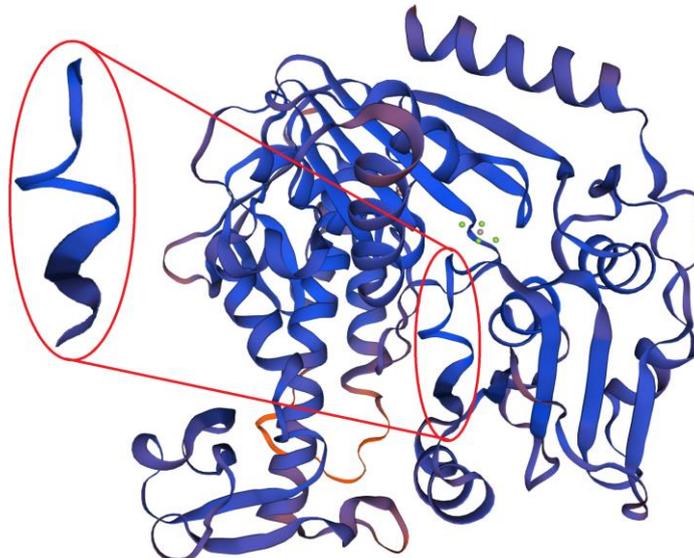


Figura 7: WNV - NSP5 - BarCODE 1

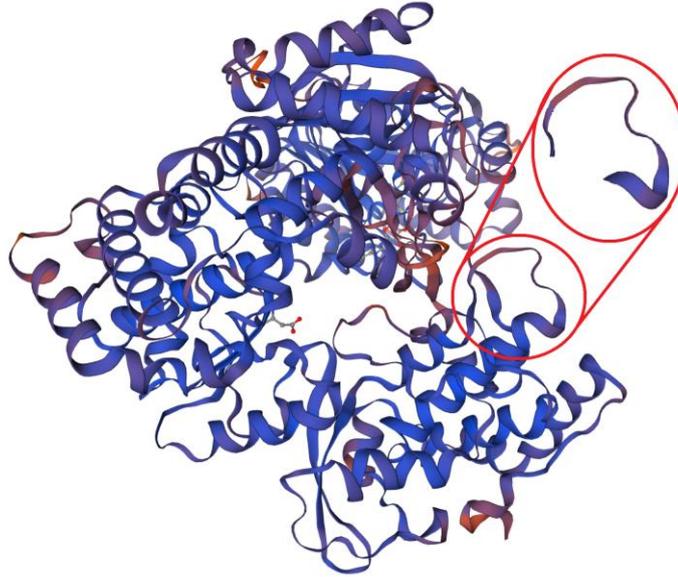


Figura 8: YFV - NSP5 - BarCODE 2

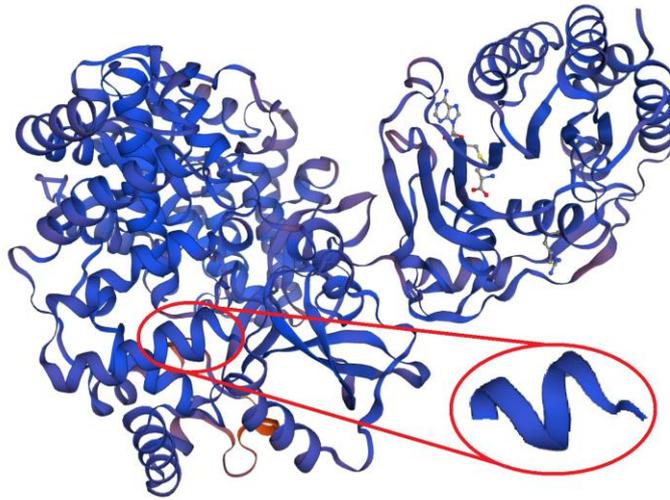
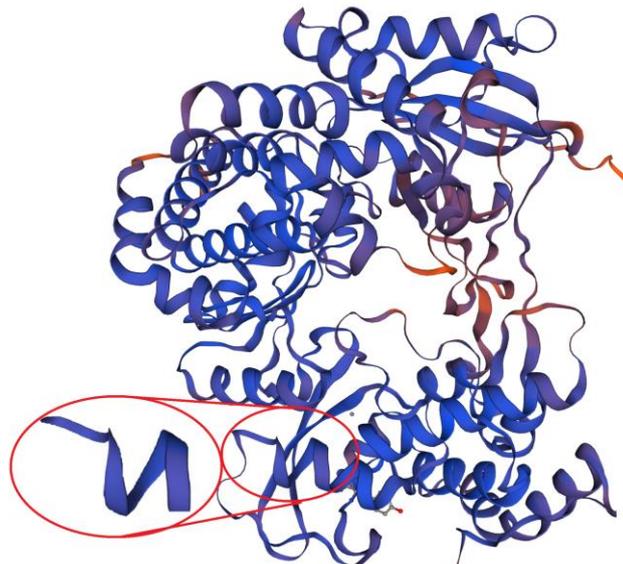


Figura 9: ZIKV - NSP5 - BarCODE 2



Os códigos de barras identificados, como apresentado e justificado previamente, ficaram menores do que o padrão, afinal, o comprimento de sequências em torno de 500 pb foi definido para espécies animais e pode ser um fator desfavorável no processo de identificação de códigos de barras virais. Dessa forma, as sequências curtas em regiões pouco mutáveis parecem ser mais apropriadas para os organismos em questão neste trabalho. Afinal, foram observados valores elevados tanto de cobertura quanto de identidade que indicam a eficiência dessas sequências no processo de identificação.

## **5 CONSIDERAÇÕES FINAIS**

Foi possível propor códigos de barras para todos os grupos virais com altos valores de confiabilidade. Cada um dos grupos virais apresentou, pelo menos, dois diferentes códigos de barras de identificação. Esses marcadores possibilitam a identificação imediata após sequenciamento de quaisquer um dos vírus analisados. Além disso, uma taxa elevada de precisão, acurácia e especificidade, quando comparados com os métodos convencionais de identificação. Assim como, conseqüentemente, fazer parte do crescimento do sistema de códigos de barras para comunidades a fim de que este deixe de ser um tópico estagnado no desenvolvimento científico.

Ainda faz-se extremamente necessária a expansão do conhecimento a respeito de códigos de barras de comunidades virais para que, além dos grupos aqui analisados, sejam determinados marcadores para diversos outros. Para isso, é preciso definir quais são os fatores específicos necessários para a definição de marcadores de identificação. Estudos futuros poderiam avaliar de forma mais detalhada os fatores que influenciam a estabilidade genética viral para definir os genes potenciais de identificação com maior facilidade, por exemplo.

O desenvolvimento cada vez mais impactante de técnicas de NGS que podem ser usadas em conjunto com os códigos de barras aqui propostos, por exemplo, pode ser um forte aliado na identificação extremamente rápida e precisa desses patógenos. Utilizando-se de corretas aplicações podem-se gerar impactos positivos em diversos setores, desde a epidemiologia e taxonomia até a própria saúde pública.

## REFERÊNCIAS

- ACOSTA-AMPUDIA, Y. *et al.* Mayaro: an emerging viral threat?. **Emerging microbes & infections**, v. 7, n. 1, p. 1-11, 2018.
- ALLAIRE, J. RStudio: integrated development environment for R. **Boston, MA**, v. 770, p. 394, 2012.
- ANDRADE, E. S. **Taxas de mutação de 14STRs autossômicas na população de Pernambuco**. 2008. 173 f. Dissertação (Mestrado em Genética) - Programa de Pós Graduação, Universidade Federal de Pernambuco Recife.
- AZEVEDO, D. A. A. *et al.* Comparação de testes sorológicos e molecular para diagnóstico da Artrite Encefalite Caprina e avaliação clínica da glândula mamária de caprinos leiteiros infectados. **Acta Scientiae Veterinariae**. v. 47, n. 1668, p. 1-8, 2019.
- BARTOLINI, I.; *et al.* Towards the implementation of a DNA barcode library for the identification of Peruvian species of *Anastrepha* (Diptera: Tephritidae). **Plos one**, v. 15, n. 1, p. e0228136, jan, 2020.
- BONALDO, M. C. *et al.* Genome analysis of yellow fever virus of the ongoing outbreak in Brazil reveals polymorphisms. **Memórias do Instituto Oswaldo Cruz**, v. 112, n. 6, p. 447–451, 2017.
- BORGHETTI, I. A. **Desenvolvimento de um controle externo positivo do tipo viral-like-particle para utilização em um teste diagnóstico de arboviroses emergentes**. 2018. 123 f. Tese (Doutorado em Engenharia de Bioprocessos e Biotecnologia) - Setor de Tecnologia, Universidade Federal do Paraná, Curitiba.
- BROWN, J. E. *et al.* Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. **Evolution**, v. 68, n. 2, p. 514-525, 2013.
- CAMPILLO-BALDERAS, J. A.; LAZCANO, A.; BECERRA, A. Viral genome size distribution does not correlate with the antiquity of the host lineages. **Frontiers in Ecology and Evolution**, v. 3, p. 143, 2015.
- CAO, C. *et al.* Detection of avian influenza virus by fluorescent DNA barcode-based immunoassay with sensitivity comparable to PCR. **Analyst**. v. 135, p. 337-342. 2010.
- CHAUDHARY, D. K.; DAHAL, R. H. DNA Bar-Code for Identification of Microbial Communities: A Mini-Review. **EC Microbiol**, v. 6, p. 219-224, 2017.
- COISSAC, E. *et al.* From barcodes to genomes: extending the concept of DNA barcoding. **Mol Ecol**, v. 25, p. 1423-1428, 2016.

- COLPITTS, T. M. *et al.* West Nile Virus: biology, transmission, and human infection. **Clinical microbiology reviews**, v. 25, n. 4, p. 635-648, 2012.
- COSTA, Z. G. A. *et al.* Evolução histórica da vigilância epidemiológica e do controle da febre amarela no Brasil. **Revista Pan-Amazônica de Saúde**, v. 2, n. 1, p. 11-26, 2011.
- DIAS, L. B. A. *et al.* Dengue: transmissão, aspectos clínicos, diagnóstico e tratamento. **Medicina (Ribeirão Preto Online)**, v. 43, n. 2, p. 143-152, 2010.
- DONALISIO, M. R.; FREITAS, A. R. R. Chikungunya in Brazil: an emerging challenge. **Revista Brasileira de Epidemiologia**, v. 18, n. 1, p. 283–305, 2015.
- DUAN, W. *et al.* The crystal structure of Zika virus NS 5 reveals conserved drug targets. **The EMBO journal**, v. 36, n. 7, p. 919-933, 2017.
- DUBANKOVA, A; BOURA, E. Structure of the yellow fever NS5 protein reveals conserved drug targets shared among flaviviruses. **Antiviral research**, v. 169, p. 104536, 2019.
- ESTRELA, J. F. **Estrutura e Patogênese das Principais Arboviroses Humanas no Brasil**. 2017. 18 f. Trabalho de conclusão de curso - Centro Universitário de Brasília (UniCEUB), Brasília.
- FERREIRA, I. R. **Aplicação do sequenciamento de nova geração no diagnóstico molecular da síndrome de Prader-Willi**. 2019.113 f. Dissertação (Mestrado em Ciências) - Pós-Graduação em Pesquisa Aplicada à Saúde da Criança e da Mulher, Fundação Oswaldo Cruz, Rio de Janeiro.
- FIGUEIREDO, M. L. G.; FIGUEIREDO, L. T. M. Emerging alphaviruses in the Americas: Chikungunya and Mayaro. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 47, n. 6, p. 677-683, 2014.
- FRIEDRICH-JANICKE, B. *et al.* Genome analysis of Mayaro virus imported to Germany from French Guiana. **Emerging infectious diseases**, v. 20, n.7, p. 1255-1257, 2014.
- FUMAGALLI, M. J. *et al.* Development of an enzyme-linked immunosorbent assay to detect antibodies targeting recombinant envelope protein 2 of Mayaro virus. **Journal of clinical microbiology**, v. 57, n. 5, 2019.
- GARCIA, A. C. H. **Métodos de Identificação Molecular de Doenças infectocontagiosas**. 2018a. 32 f. Trabalho de conclusão de curso - Centro Universitário de Brasília (UniCEUB), Brasília.
- GARCIA, L. P. Epidemia do vírus Zika e microcefalia no Brasil: emergência, evolução e enfrentamento. Texto para Discussão. **Instituto de Pesquisa Econômica Aplicada - IPEA**. Brasília, 2018b.

- GEBHARD, L. G.; FILOMATORI, C. V.; GAMARNIK, A. V. Functional RNA elements in the dengue virus genome. **Viruses**, vol. 3, n. 9, p. 1739-1756. 2011.
- GROCHOCKI, T. M. **Principais métodos de detecção de viroses transmitidas por Aedes no Brasil**. 2016. 31 f. Trabalho de Conclusão de Curso - Centro Universitário de Brasília (UniCEUB), Brasília.
- GUBLER, D. J.; VASILAKIS, N.; MUSSO, D. History and emergence of Zika virus. **Journal of Infectious Diseases**, v. 216, p. 860-867, 2017.
- GUIMARÃES, M. **Desenho racional de antivirais: Identificação de Novos Inibidores Contra a Protease do Vírus do Nilo Ocidental por High-Throughput Screening**. 2021. Dissertação (Mestrado em Biotecnologia) - Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo, São José dos Campos.
- HADDOW, A. D. et al. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. **PLoS neglected tropical diseases**, v. 6, n. 2, e1477, 2012.
- HALL, T.; BIOSCIENCES, I.; CARLSBAD, C. BioEdit: an important software for molecular biology. **GERF Bull Biosci**, v. 2, n. 1, p. 60-61, 2011.
- HAJIBABAEI, M. *et al.* **Methods in Molecular Biology**. Totowa, NJ. Humana Press, 2012. p. 339-353.
- HAMMER, Ø.; HARPER, D. A. T.; RYAN, P. D. PAST: Paleontological statistics software package for education and data analysis. **Palaeontologia electronica**, v. 4, n. 1, p. 9, 2001.
- HEBERT, P. D. N. *et al.* Biological identifications through DNA barcodes. **Proceedings of the Royal Society of London. Series B: Biological Sciences**, v. 270, n. 1512, p. 313-321, 2003.
- HOLMES, E. C.; TWIDDY, S. S. The origin, emergence and evolutionary genetics of dengue virus. **Infection, Genetics and Evolution**, v. 3, n.1, p. 19–28. 2003.
- HONÓRIO, N. A. *et al.* Chikungunya: an arbovirus infection in the process of establishment and expansion in Brazil. **Cad. Saúde Pública**, vol. 31, n. 5, p. 1-3, 2015.
- KEARSE, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647-1649, 2012.
- KENNEDY, M. A.; PARKS, R. J. Adenovirus virion stability and the viral genome: size matters. **Molecular Therapy**, v. 17, n. 10, p. 1664-1666, 2009.

- KERKHOF, K. *et al.* Reliable Serological Diagnostic Tests for Arboviruses: Feasible or Utopia? **Trends in Microbiology**, v. 28, n. 4, p. 276-292, 2020.
- KLITTING, R. *et al.* Molecular determinants of Yellow Fever Virus pathogenicity in Syrian Golden Hamsters: one mutation away from virulence. **Emerging microbes & infections**, v.7, n.1, p. 1-18. 2018.
- KINDHAUSER, M. K. *et al.* Zika: the origin and spread of a mosquito-borne virus. **Bull World Health Organ**, vol. 94, p. 675-686, 2016.
- KRAEMER, M. U. G. *et al.* The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. **eLife**, v. 4, p. 1-18, 2015.
- LEITE, A. A.; ERRANTE, P. R. Asp ctos cl nicos, preven o e epidemiologia da febre amarela no Brasil. **UNILUS**. v. 14, n. 34, p. 169-184, 2017.
- LIMA-CAMARA, T. N. Emerging arboviruses and public health challenges in Brazil. **Revista de Sa de P blica**, v. 50, n. 36, p. 1-7, 2016.
- M LAGA, D. E. R. **Avalia o de valida o da utilidade cl nica do sequenciamento de nova gera o (NGS) para confirma o do diagn stico de doen as lisoss micas selecionadas**. 2018. 157 f. Tese (Doutorado em Gen tica e Biologia Molecular) - Programa de P s-Gradua o, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- MAAN, S. *etal.* Complete Genome Characterisation of a Novel 26th Bluetongue Virus Serotype from Kuwait. **PLoS One**. v. 6, n. 10, e26147, 2011.
- MASCARENHAS, M rcio D nis Medeiros et al. Ocorr ncia simult nea de COVID-19 e dengue: o que os dados revelam? **Cadernos de Sa de P blica**, v. 36, p. e00126520, 2020.
- May, F. J. *et al.* Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. **Journal of virology**, v. 85, n. 6, p. 2964-2974, 2011.
- MEDEIROS, I. E. A. F.; BRAGA, G. M. S. Efeitos sociais provocados pelo n mero de casos de doen as transmitidas pelo mosquito *Aedes aegypti* no munic pio de Balsas. **PUBVET**, v. 12, n.10, a193, 1-5, out, 2018.
- MESSINA, J. P. et al. Global spread of dengue virus types:mapping the 70 year history. **Trends in Microbiology**, v. 22, n. 3, 2014.
- MINIST RIO DA SA DE. Monitoramento dos casos de arboviroses urbanas transmitidas pelo *Aedes Aegypti* (dengue, chikungunya e zika), Semanas Epidemiol gicas 1 a 32, 2020. **Boletim Epidemiol gico**, v. 51, n. 16, p. 11-17, 2020.

- MINISTÉRIO DA SAÚDE. Monitoramento dos casos de arboviroses urbanas transmitidas pelo *Aedes Aegypti* (dengue, chikungunya e zika), Semanas Epidemiológicas 1 a 21, 2021. **Boletim Epidemiológico**, v. 52, n. 21, p. 1-11, 2021.
- MELO, S. F. **Avaliação de citocinas IL-17 e IL-23 no soro de pacientes com dengue na cidade de Manaus**. 2014. 19 f. Relatório Final PIB - Programa Institucional de Iniciação Científica, Universidade Federal do Amazonas, Manaus.
- MIR, D. *et al.* Phylodynamics of yellow fever virus in the Americas: new insights into the origin of the 2017 Brazilian outbreak. **Scientific reports**, v. 7, n.1, p. 1-9. 2017.
- MONCAYO, A. C. *et al.* Dengue emergence and adaptation to peridomestic mosquitoes. **Emerging Infectious Diseases**. v. 10, n. 10, p. 1790–1796, 2004.
- MURRAY, K. O. West Nile Virus and its emergence in the United States of America. **VBet Res.** v. 41, n. 6, p. 67, 2010.
- MOTA, M. T. O. *et al.* Mayaro virus: a neglected arbovirus of the Americas. **Future virology**, v. 10, n. 9, p. 1109-1122, 2015.
- MUSSO, D. *et al.* Detection of Zika virus in saliva. **Journal of Clinical Virology**, v. 68, p. 53-55, 2015.
- MUSTAFA, M. S. *et al.* Discovery of fifth serotype of dengue virus (DENV-5): A new public health dilemma in dengue control. **Medical Journal Armed Forces India**, v. 71, n. 1, p. 67-70, 2015.
- NASCIMENTO, J. R. **Distribuição espacial de infecção humana por Dengue, Chikungunya e Zika**. 2018. Dissertação (Mestrado em Entomologia e Conservação da Biodiversidade) - Programa de Pós-Graduação, Universidade Federal da Grande Dourado, Dourados.
- NORONHA, T. G.; CAMACHO, L. A. B. Controvérsias sobre a ampliação das áreas com vacinação de rotina contra a febre amarela no Brasil. **Cad. Saúde Pública**. v. 33, n. 10, p. e00060917, 2017.
- OLIVEIRA, R. M. B. **Produção de imunoglobulina humana: nova aplicação terapêutica para o tratamento da dengue**. 2012. Dissertação (Mestrado em Inovação Terapêutica) - Programa de Pós-Graduação, Universidade Federal de Pernambuco, Recife.
- OLIVEIRA, M. M. **Revisão integrativa sobre os testes diagnósticos da febre de Chikungunya**. 2018. 60f. Trabalho de Conclusão de Curso - Fundação Oswaldo Cruz, Rio de Janeiro.
- OLIVEIRA, A. S. *et al.* NS3 and NS5 proteins: important targets for anti-dengue drug design. **Journal of the Brazilian Chemical Society**, v. 25, n. 10, p. 1759-1769, 2014.

- OMETTO, T. L. **Monitoramento do vírus do Oeste do Nilo no Brasil**. 2013. Tese (Doutorado em Ciências) - Programa de Pós-Graduação USP / IPT / Instituto Butantan, São Paulo.
- PAUL, A.; VIBHUTI, A. Dengue Symptoms Significance in Anti-Dengue Drug Development: Road Less Travelled. **Bioinformatics**, vol. 13, n. 5, p. 131-135, 2017.
- PESENTI, A. L.; SILVA, E. Vírus zika–Epidemiologia e diagnóstico laboratorial. **Rev Med Minas Gerais**, v. 28, p. 1-5, 2018.
- PINHO, A. C. O. **Diagnóstico e caracterização molecular do vírus dengue circulante na cidade de Salvador, Bahia, Brasil**. 2013. 75 f. Dissertação (Mestrado em Biotecnologia) - Programa de Pós-Graduação em Biotecnologia do Instituto de Ciências da Saúde, Universidade Federal da Bahia, Salvador.
- PINTO, L. B. **Investigação dos vírus da influenza, doença de newcastle e vírus do nilo ocidental em aves no pantanal matogrossense**. 65 f. Dissertação (Mestrado) - Universidade Federal do Mato Grosso.
- PINTO-JUNIOR, V. L.; LUZ, K.; PARREIRA, R.; FERRINHO, P. Vírus Zika: revisão para clínicos. **Revista Científica da Ordem dos Médicos**, v. 28, n. 6. p. 760-765, 2015.
- POWELL, J. R.; TABACHNICK, W. J. History of domestication and spread of *Aedes aegypti* - A Review. **Mem. Inst. Oswaldo Cruz**, Rio de Janeiro , v. 108, supl. 1, p. 11-17, 2013.
- PRASADINI, M. et al. Blood Feeding Preference of Female *Aedes aegypti* Mosquitoes for Human Blood Group Types and Its Impact on Their Fecundity: Implications for Vector Control. **American Journal of Entomology**. v. 3, n. 2, p. 43-48, 2019.
- PURTY, R. S.; CHATTERJEE, S. DNA barcoding: an effective technique in molecular taxonomy. **Austin J Biotechnol Bioeng**, v. 3, n. 1, 1059, p. 1 - 10, 2016.
- RATNASINGHAM, Sujeevan; HEBERT, Paul DN. BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). **Molecular ecology notes**, v. 7, n. 3, p. 355-364, 2007.
- ROCHA, A. P. S. et al. A. Dengue: história natural e definição de casos graves e potencialmente graves. **Rev Med Minas Gerais**, v. 21, n. 4, Supl 6, p. 49-53, 2011.
- ROGERS, D. J. *et al.* The global distribution of yellow fever and dengue. **Advances in parasitology**, v. 62, p. 181-220. 2006.
- SAMPAIO, L. S. **Aplicação de DNA Barcoding na identificação de espécies dos gênero *Senna*, *Lantana* e *Casaria***. 2010. 111 f. Dissertação (Mestrado em Biotecnologia) - Instituto de Química, Universidade Estadual Paulista, Araraquara.

- SAMUEL, G. H. *et al.* Yellow fever virus capsid protein is a potent suppressor of RNA silencing that binds double-stranded RNA. **Proceedings of the National Academy of Sciences**, v. 113, n. 48, p. 13863-13868, 2016.
- SANJUAN, R. *et al.* Viral mutation rates. **Journal of virology**, v. 84, n. 19, p. 9733-9748, 2010.
- SANTOS, P. R. **Identificação de isolados de *Trichoderma spp.* utilizando marcadores do tipo RAPD e DNA BarCODE.** 2010. 69 f. Dissertação (Mestrado Biologia Celular e Molecular) - Programa de Pós-Graduação, Universidade Federal de Goiás, Goiânia.
- SILVA, J. W. P. **Aspectos ecológicos de vetores putativos do Vírus Mayaro e Vírus Oropuche em estratificação vertical e horizontal em ambientes florestais e antropizados em uma comunidade rural no Amazonas.** 2017. 97 f. Dissertação (Mestrado em Saúde Coletiva) - Instituto Leônidas e Maria Deane, Fundação Oswaldo Cruz, Manaus.
- SILVA, J. W. P. Doença pelo vírus do Oeste do Nilo: um novo problema para o Brasil. **Epidemiol. Serv. Saúde**, v. 25, n. 3, 2016.
- SCHUFFENECKER, I. *et al.* Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. **PLoS Med**, v. 3, n. 7, p. e263, 2006.
- SOLIGNAT, M. *et al.* Replication cycle of chikungunya: a re-emerging arbovirus. **Virology**, v. 393, n. 2, p. 183-197, 2009.
- SONG, B. H. *et al.* Zika virus: history, epidemiology, transmission, and clinical presentation. **Journal of neuroimmunology**, v. 308, p. 50-64, 2017.
- SONI, M. *et al.* DNA barcoding as a complementary approach for species identification from dengue endemic regions of North East India. **International Journal of Mosquito Research.**; v. 5, n. 1, p. 46-52, 2018.
- SOUZA, A. R. V. **Avaliação sorológica e molecular de pacientes com quadro clínico de dengue símile e atendidos no hospital das forças armadas.** 2019. 116 f. Dissertação (Mestrado em Genética e Biologia Molecular) - Programa de Pós-Graduação, Universidade Federal do Goiás, Goiânia.
- STROHER, P. R. Sequenciamento de nova geração e entomologia: Novas perspectivas para antigos questionamentos. **Revista da Biologia.** v. 18, n. 1, p. 6-16, 2018.
- SUCUPIRA, P. H. F. **Avaliação da resposta imune do vetor *Aedes aegypti* à infecção por Mayaro vírus.** 2019. 92 f. Dissertação (Mestrado em Biologia Celular e Molecular, Genética e

Bioinformática) - Programa de Pósgraduação em Ciências da Saúde do Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, 2019.

TEAM, R. Core *et al.* R: A language and environment for statistical computing. Vienna, Austria. 2013.

TEICH, V.; ARINELLI, R.; FAHHAM, L. *Aedes aegypti* e sociedade: o impacto econômico das arboviroses no Brasil. **Jornal Brasileiro de Economia da Saúde**, v. 9, n. 3, p.267-276, 2017.

TERRA, M. R. *et al.* *Aedes aegypti* e as arboviroses emergentes no Brasil. **Uningá Review**, v. 30, n.3, p.52-60, 2017.

TURCHETTO, C. *et al.* Marcadores genéticos baseados em DNA. **Marcadores Moleculares na Era Genômica: Metodologias e Aplicações**. Ribeirão Preto: Sociedade Brasileira de Genética, 2017. p. 12 - 20.

VASCONCELOS, P. F. C. Febre amarela. **Rev. Soc. Bras. Med. Trop.**, Uberaba, v. 36, n. 2, p. 275-293, 2003.

VASCONCELOS, P. F. C. Doença pelo vírus Zika: um novo problema emergente nas Américas?. **Revista Pan-Amazônica de Saúde**, v. 6, n.2, p. 9-10, 2015.

VASILAKIS, N. *et al.* Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health. **Nat Rev Microbio**, v. 9, n. 7, p. 532–541, 2011.

VIANA, J. L.; *et al.* Isolates of *Bacillus thuringiensis* from Maranhão biomes with potential insecticidal action against *Aedes aegypti* larvae (Diptera, Culicidae). **Brazilian Journal of Biology**, v. 6984, p. 1–11, 2020.

VIEIRA, M. A. C. S. *et al.* West Nile virus encephalitis: the first human case recorded in Brazil. **The American journal of tropical medicine and hygiene**, v. 93, n. 2, p. 377-379, 2015.

VIEIRA, Y. P. *et al.* Descrição epidemiológica da febre amarela no Brasil: alerta sobre a expansão da doença. **Brazilian Journal of Health Review**, v. 3, n. 5, p. 13383-13395, 2020.

WEAVER, S. C.; LECUIT, M. Chikungunya Virus and the Global Spread of a Mosquito-Borne Disease. **New England Journal of Medicine**, v. 372, n. 13, p. 1231–1239, 2015.

Weeratunga, P.; *et al.* Control methods for *Aedes albopictus* and *Aedes aegypti*. **The Cochrane Database of Systematic Reviews**, v. 8, 2017.

WEI, C.; WANG, G.; CHENG, X.; HUANG, H.; LIU, B.; XU, Y.; LI, F. Identification and Typing of Human Enterovirus: A Genomic Barcode Approach. **PLoS One**. v. 6, n. 10, e26296. 2011.

WILDER-SMITH, A. GUBLER, D. J.; WEAVER, S. C.; MONATH, T. P.; HEYMANN, D. L.; SCOTT, T. W. Epidemic arboviral diseases: priorities for research and public health. **The Lancet infectious diseases**, v. 17, n. 3, p. e101-e106, 2017.

YAP, M. L.; KLOSE, T.; URAKAMI, A.; HASAN, S. S.; AKAHATA, W.; ROSSMANN, M. G. Structural studies of Chikungunya virus maturation. **PNAS**, v. 114 n. 52, p. 13703-13707, 2017.

YIN, H. *et al.* Nanoparticle-based bio-barcode assay for the detection of bluetongue virus. **Journal of Virological Methods**. v. 178, p. 225-228, 2011.

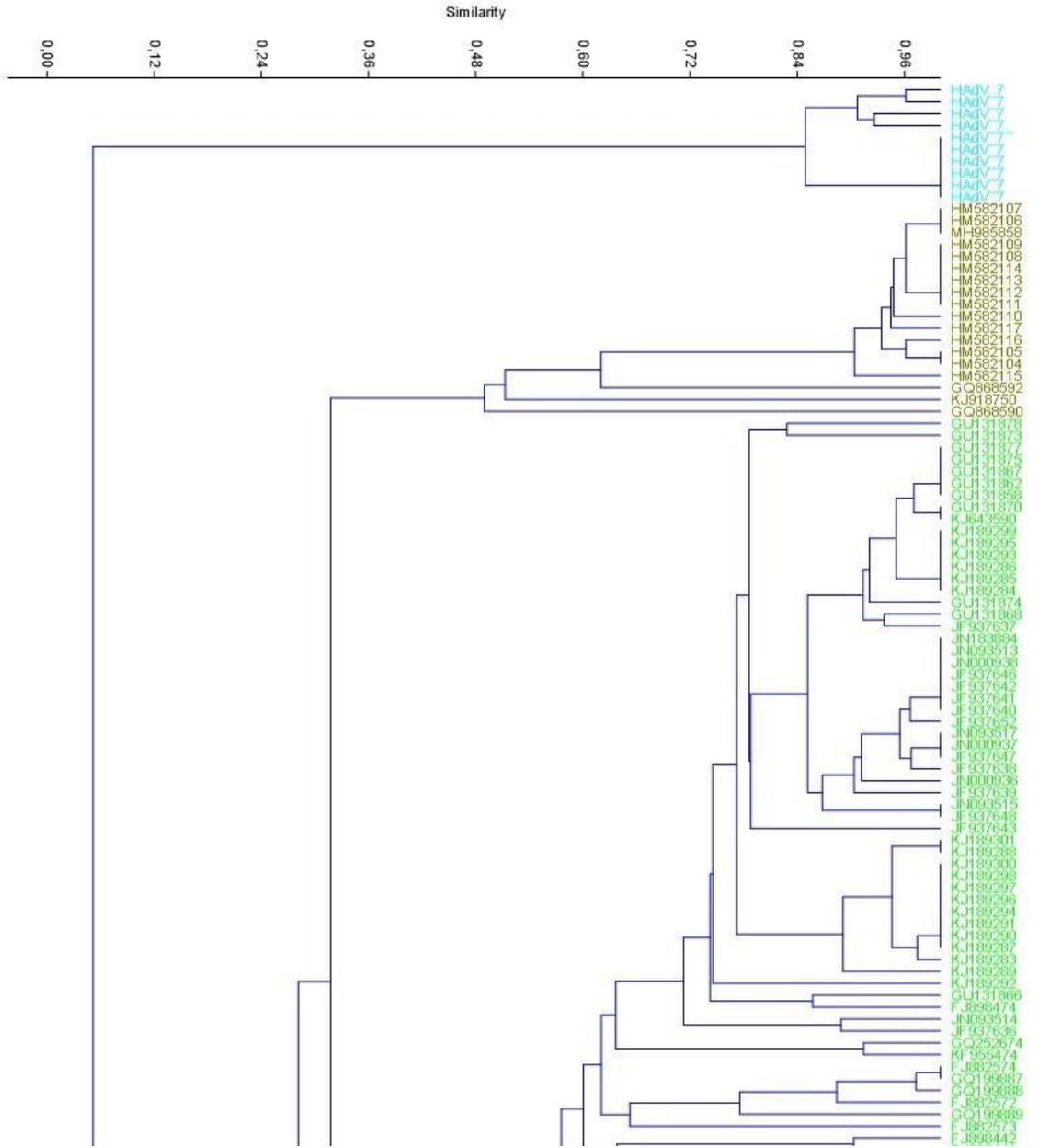
YOSHII, K. *et al.* A conserved region in the prM protein is a critical determinant in the assembly of flavivirus particles. **Journal of general virology**, v. 93, n. 1, p. 27-38, 2012.

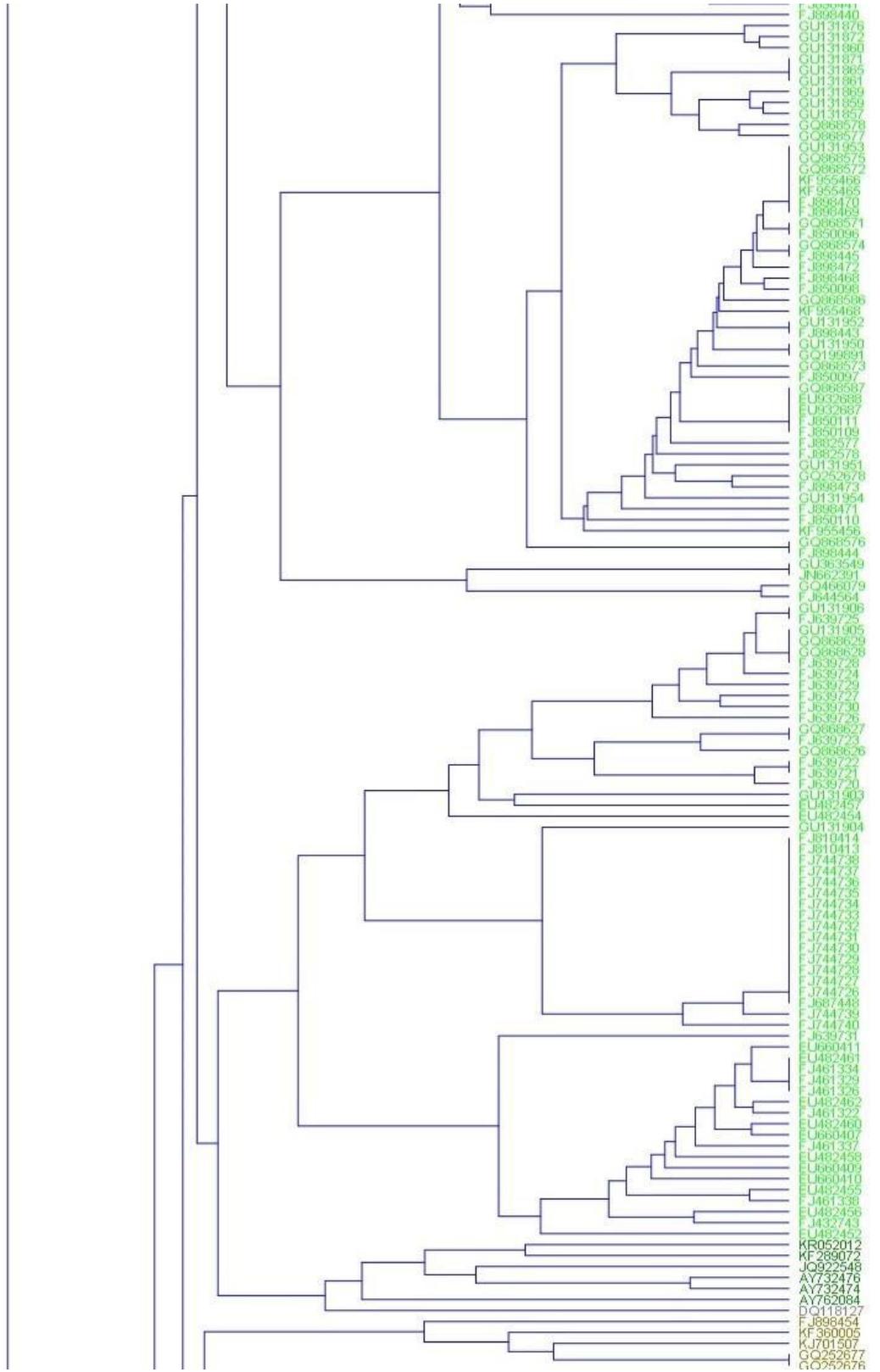
ZAMORA-DELGADO, J.; CASTANO, J. C.; HOYOS-LOPEZ, R. DNA barcode sequences used to identify *Aedes (Stegomyia) albopictus* (Diptera: Culicidae) in La Tebaida (Quindío, Colombia). **Rev. Colomb. Entomol.** v. 41, n. 2, p. 212-217, 2015 .

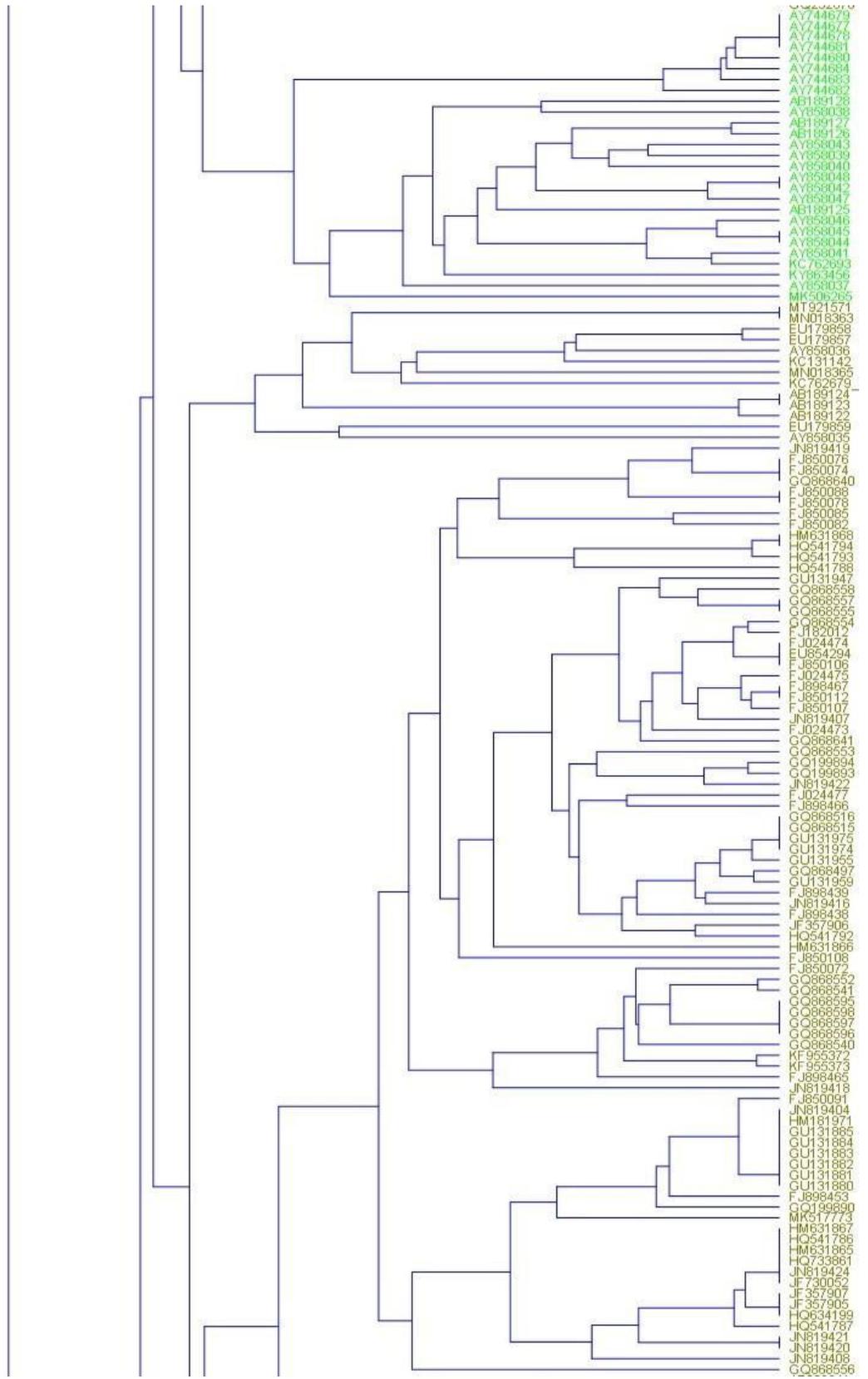
## APÊNDICES

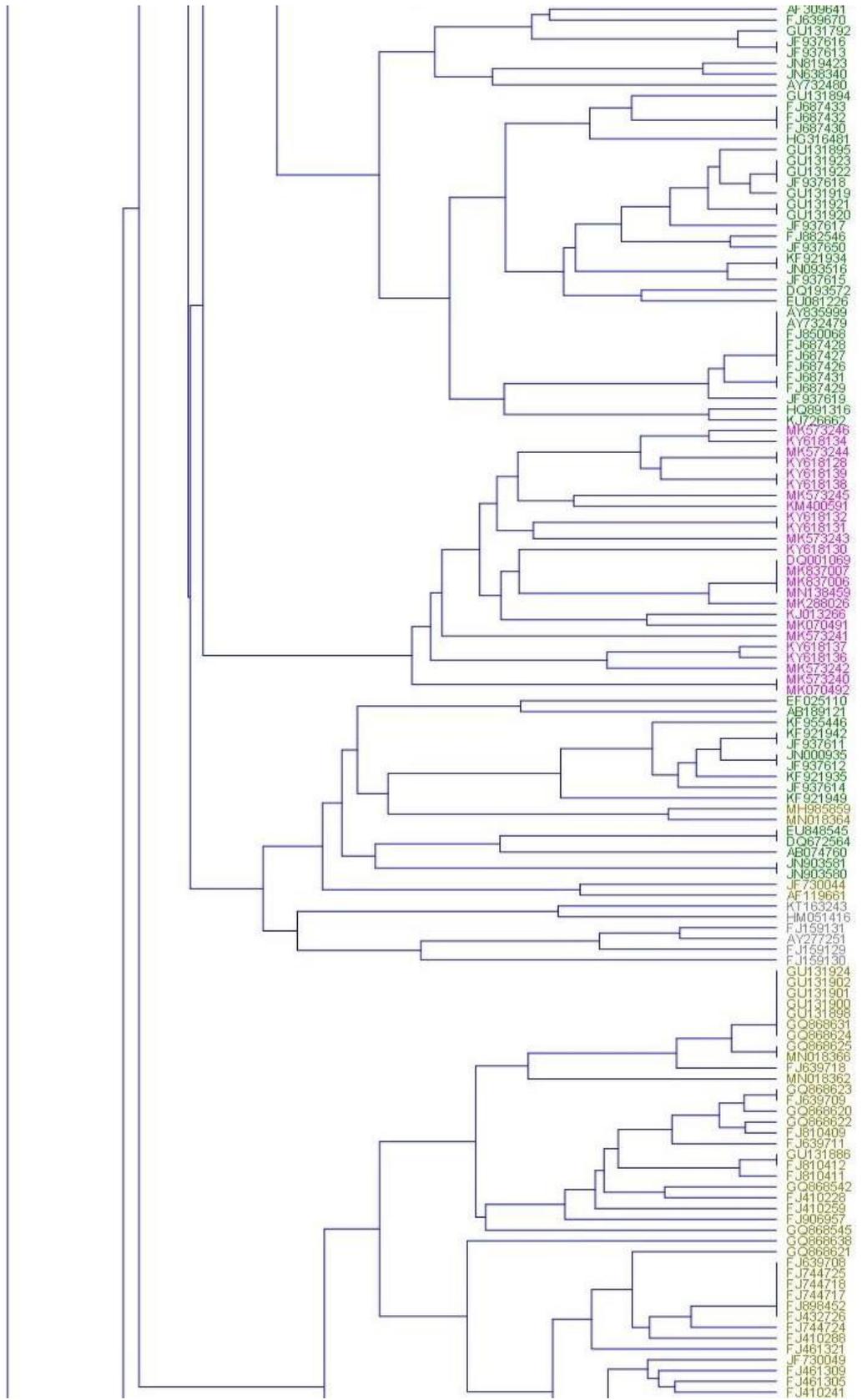
Apêndice A: Tabela completa de dados das estirpes e perfil de corte enzimático disponível em: <<https://cutt.ly/yn2XP3s>>.

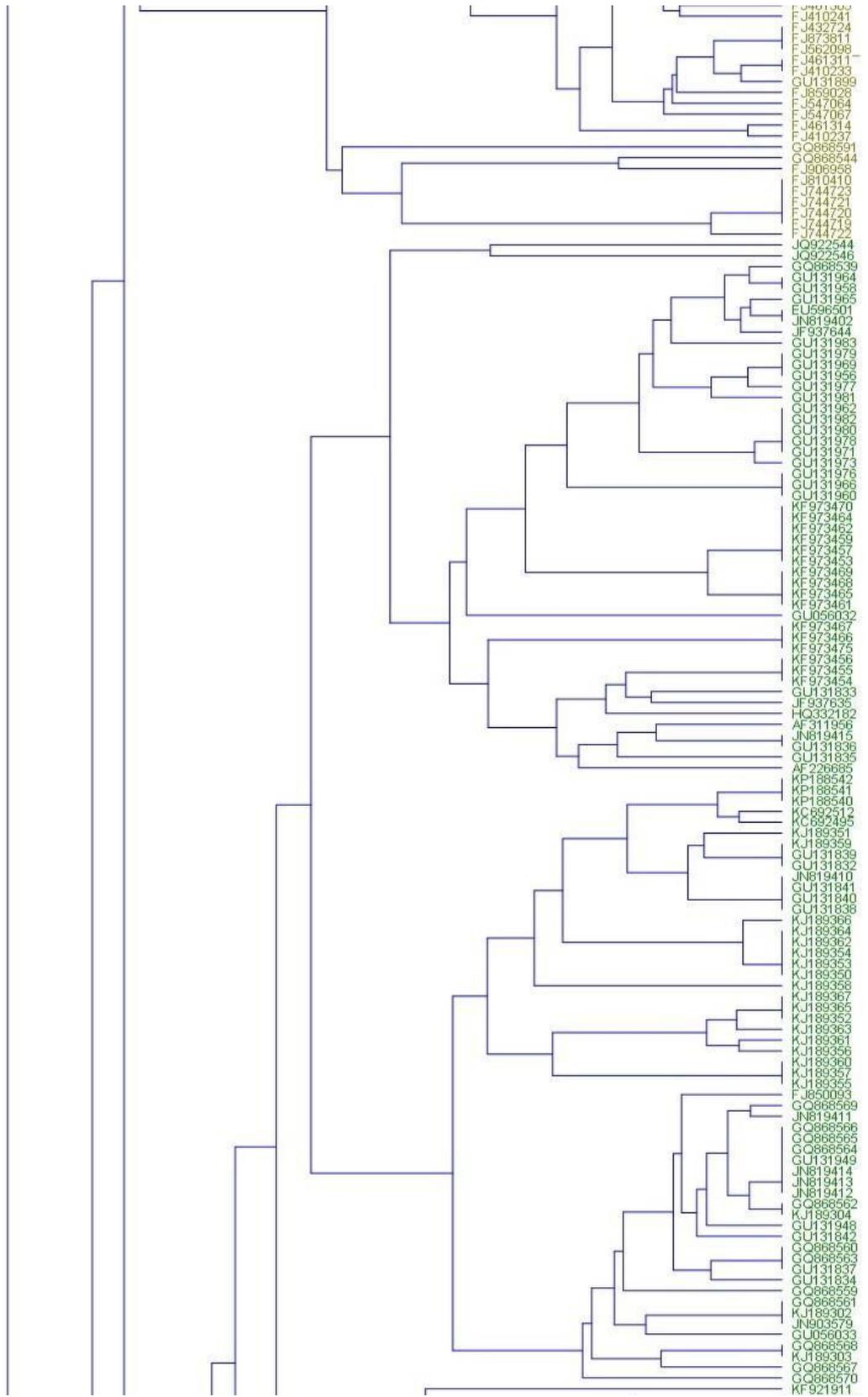
Apêndice B: Árvore filogenética completa com base no perfil de cortes enzimático das estirpes.

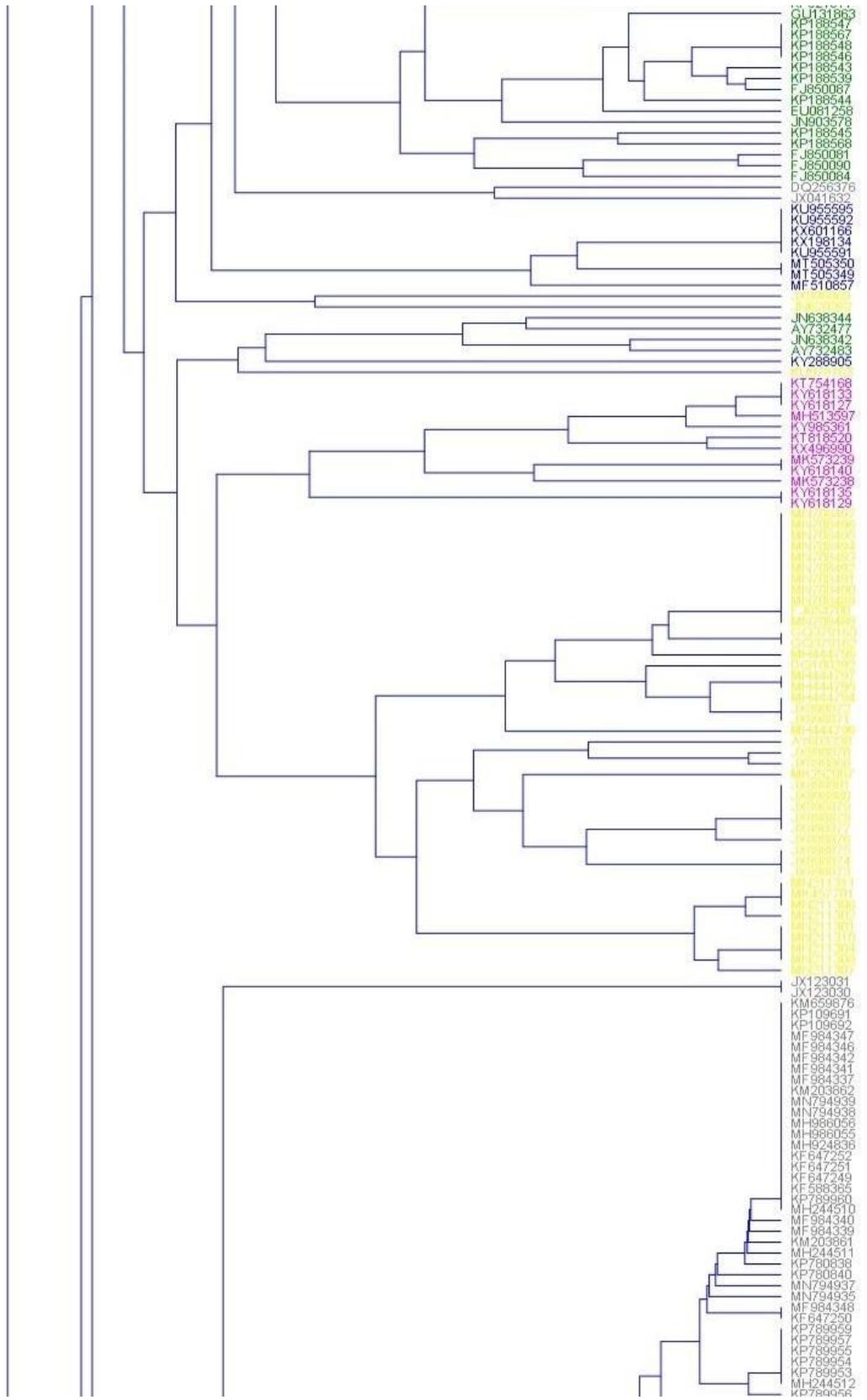


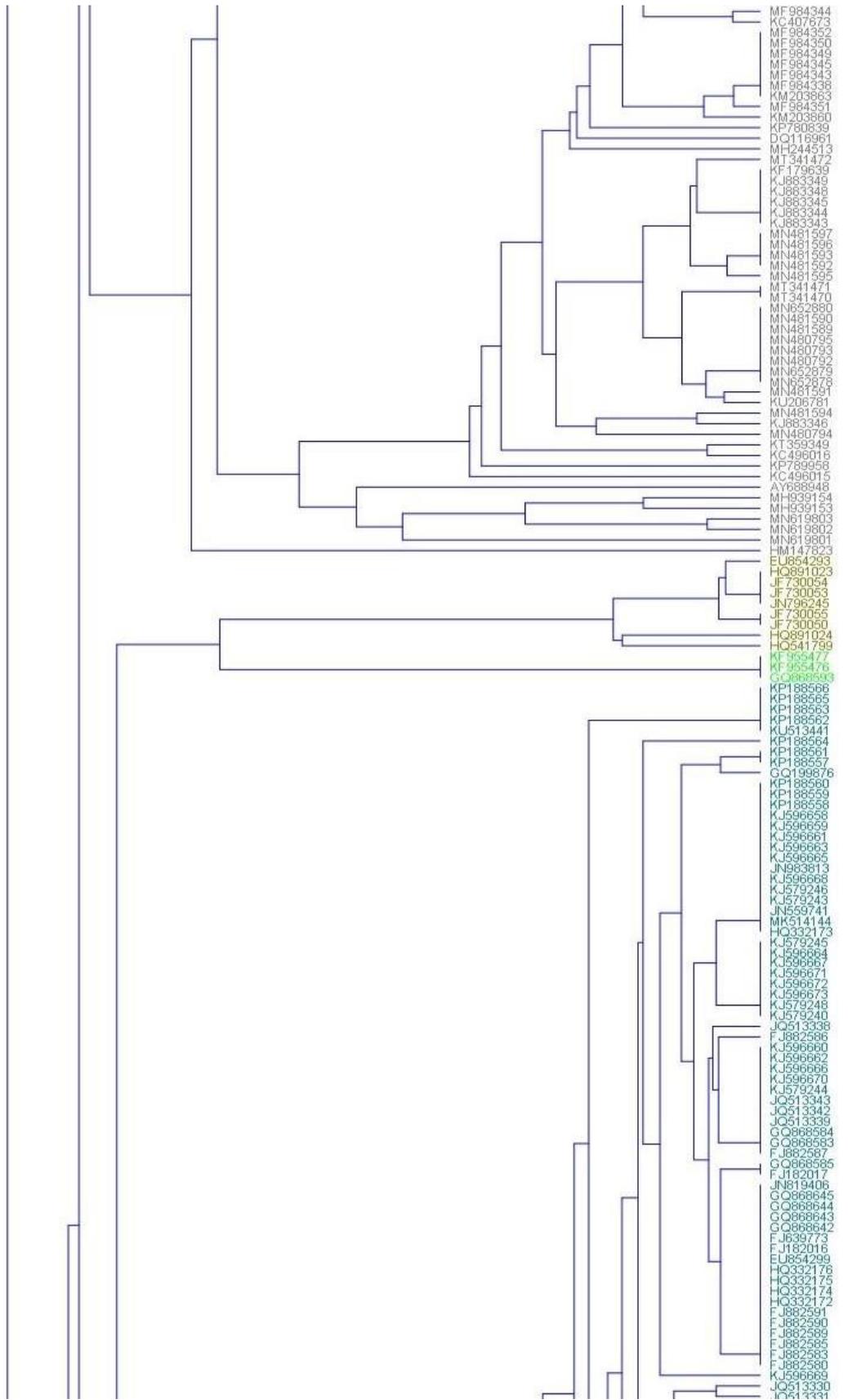


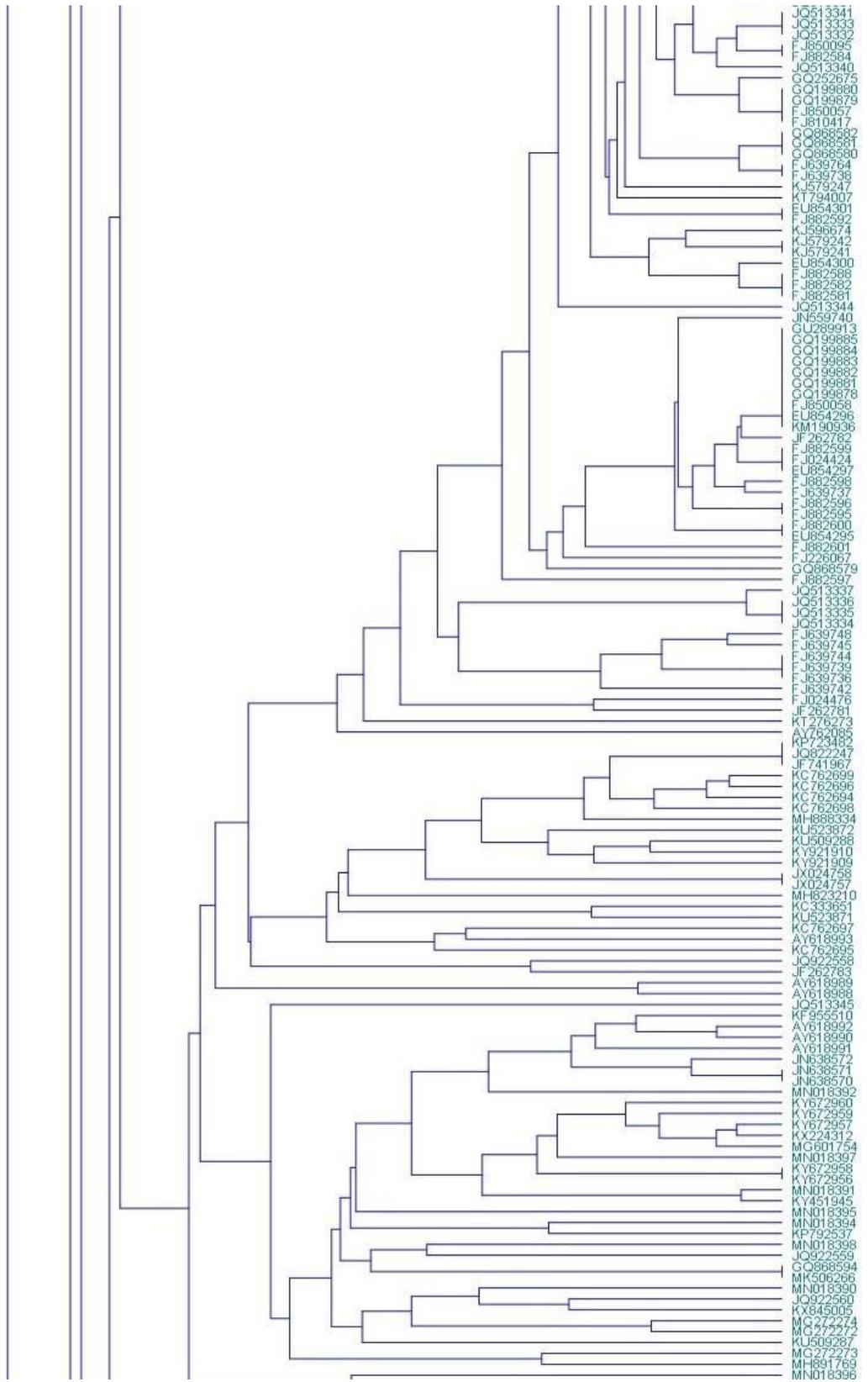


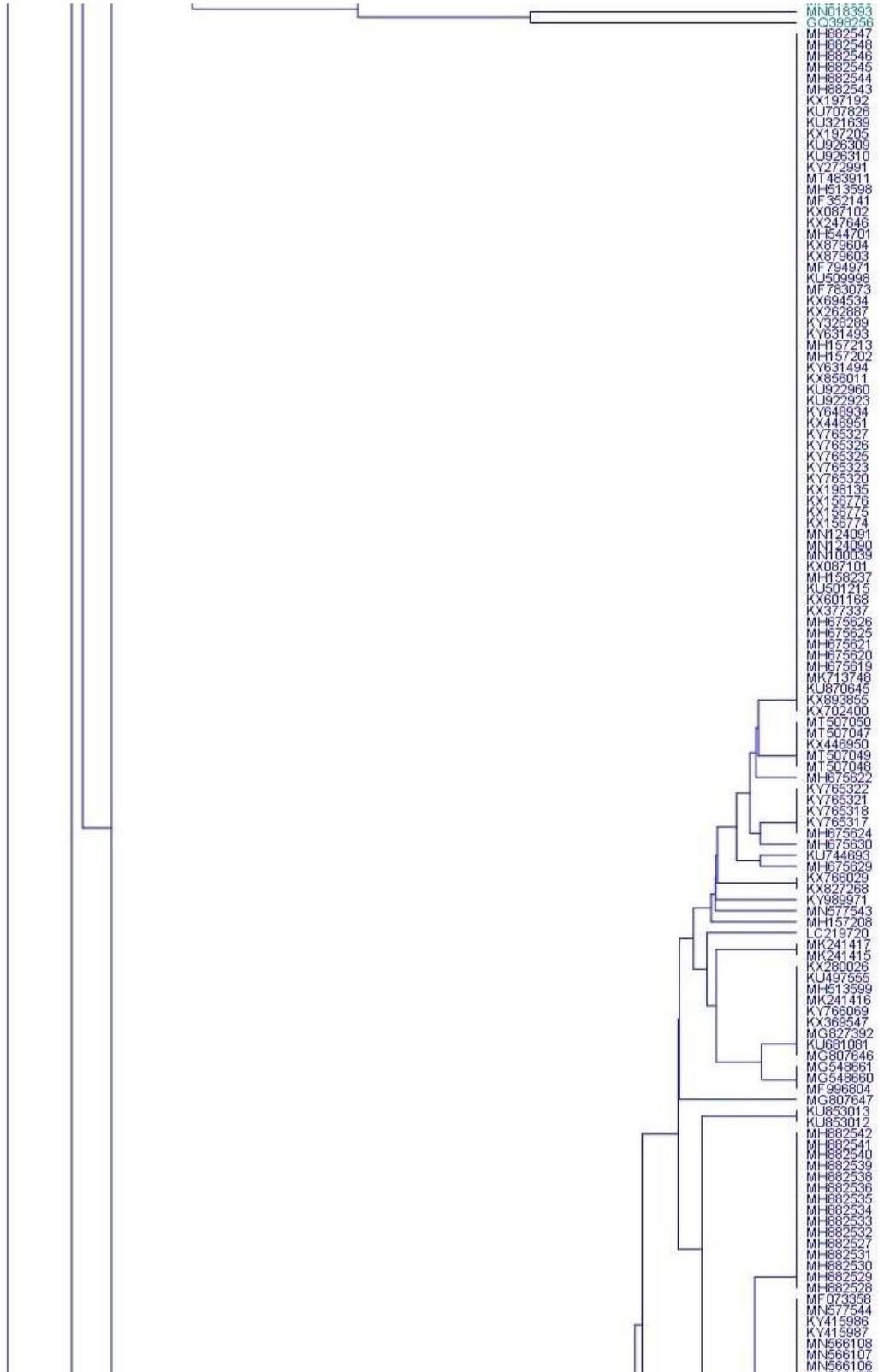


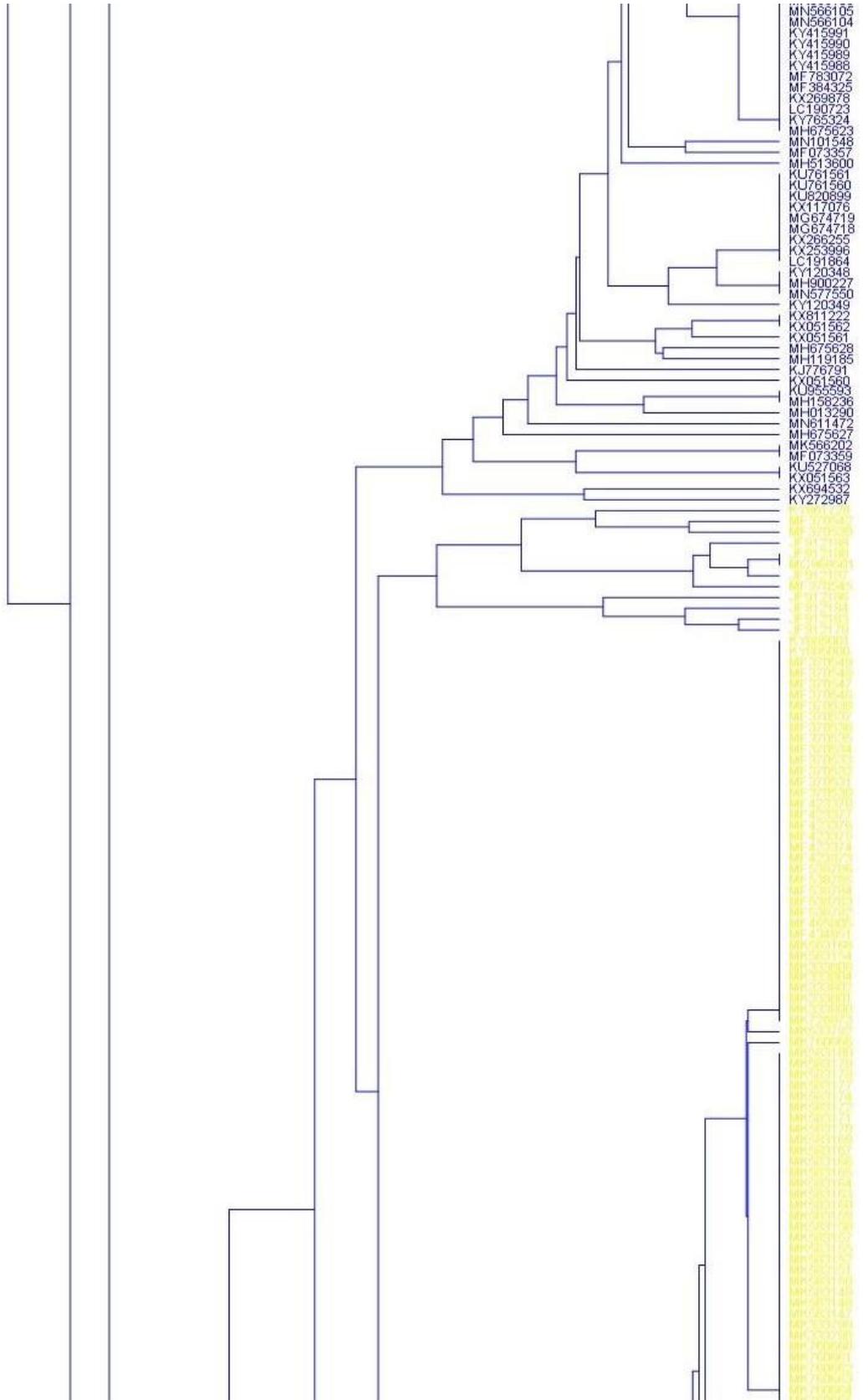


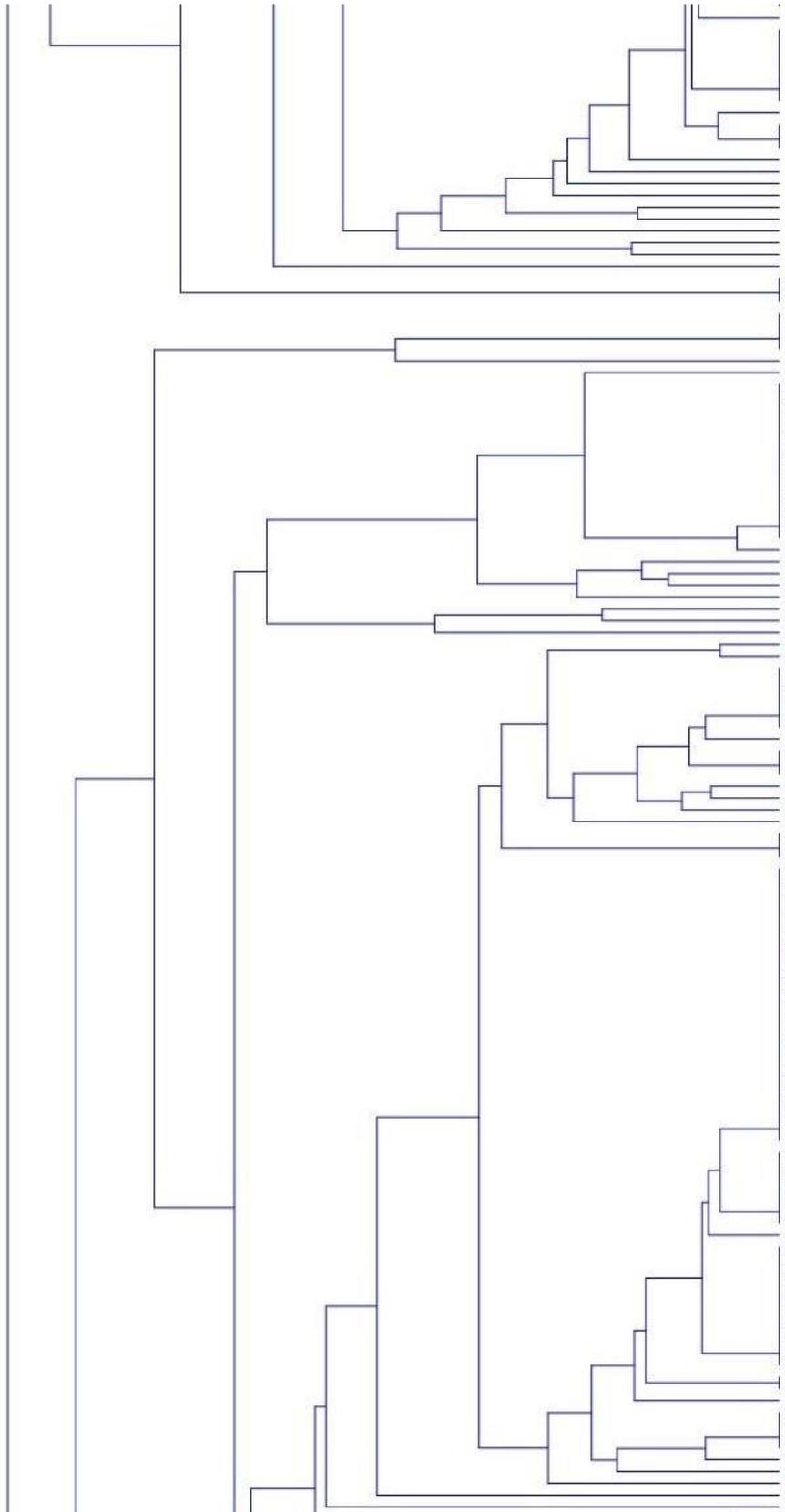




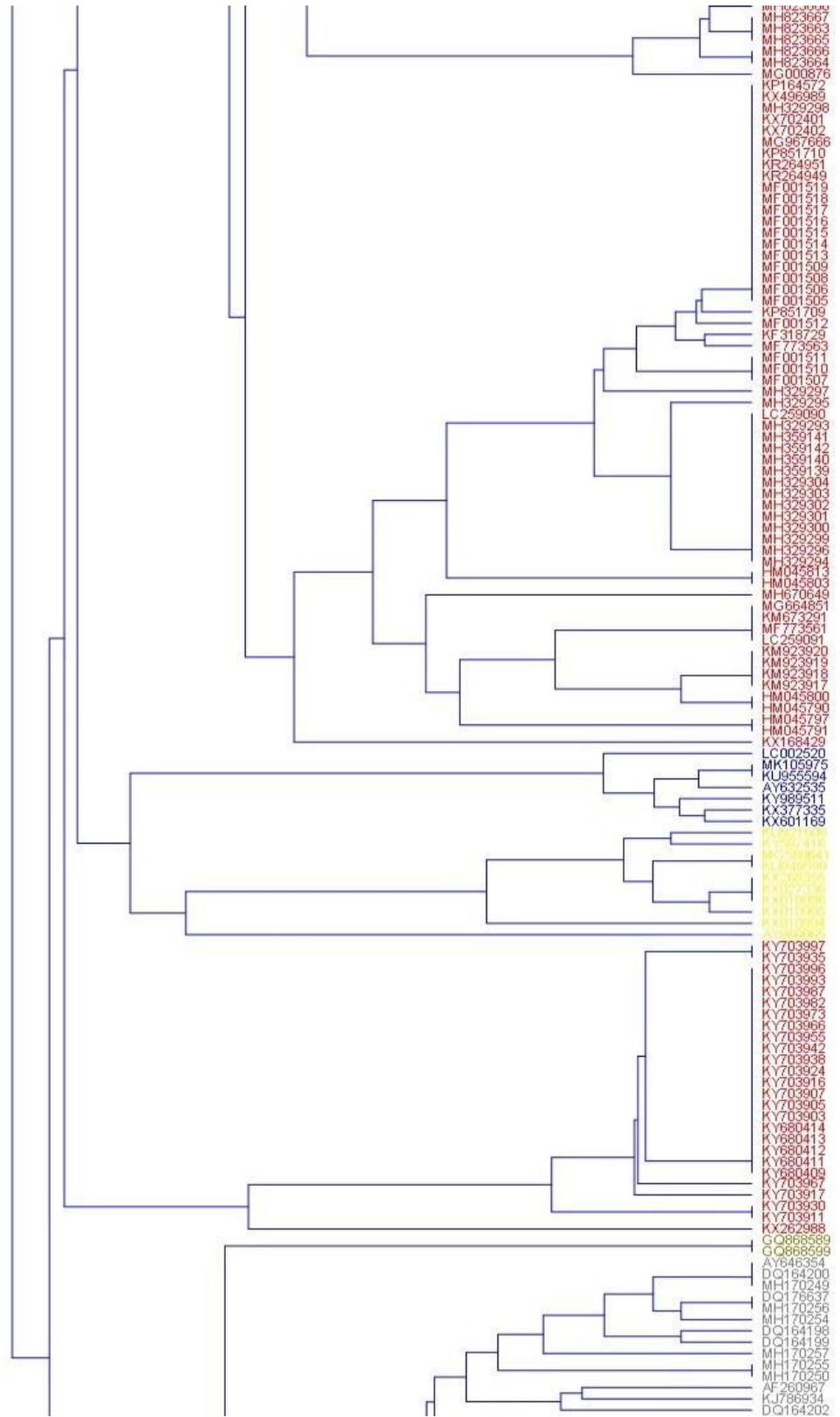


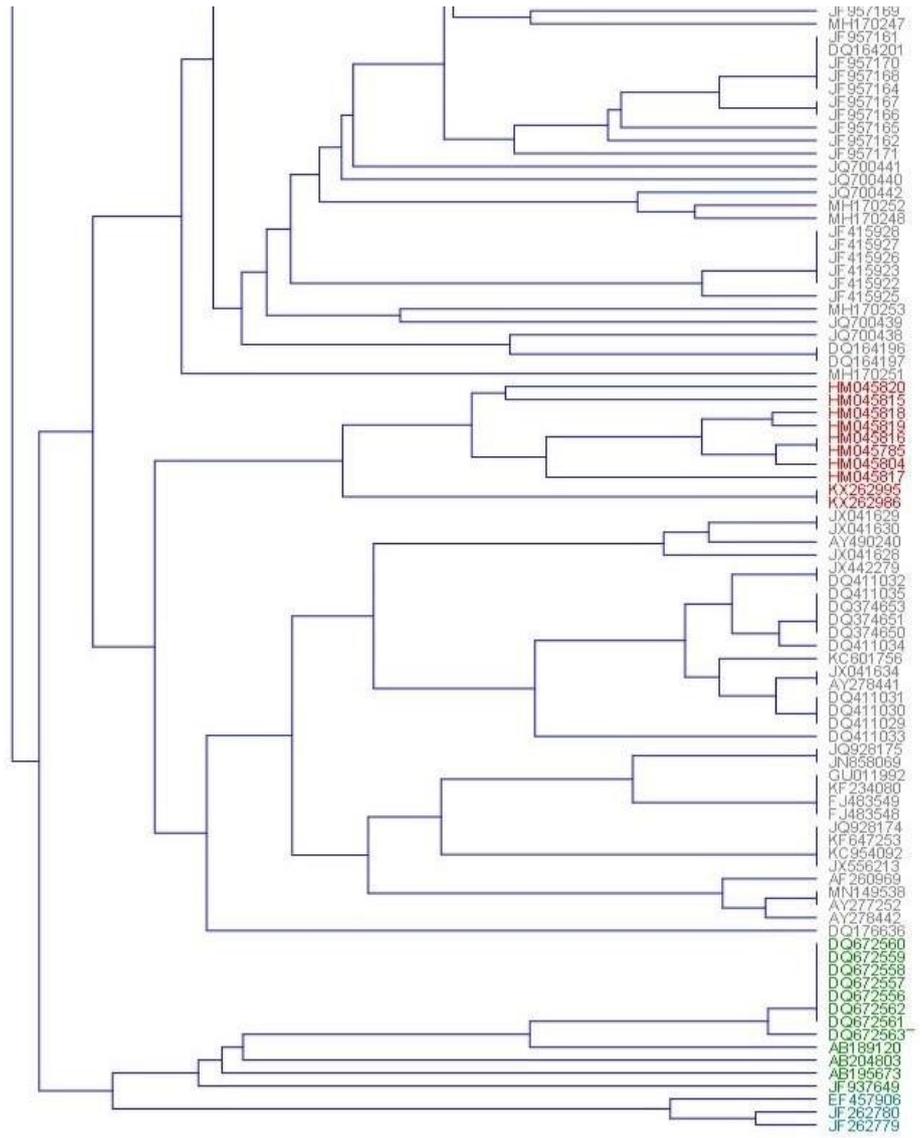






KX377336  
KX694533  
KX601167  
MT076935  
MT076934  
MT076933  
MT076932  
JQ922547  
LC259094  
MT636922  
MT636920  
MT636908  
MT636921  
MT636920  
MT636919  
MT636918  
MT636917  
MT636916  
MT636914  
MT636912  
MT636911  
MT636910  
MT636909  
MT636913  
KY038947  
KY038946  
HM045793  
HM045822  
HM045809  
HM045821  
HM045806  
HM045823  
HM045784  
JF274082  
MK120202  
MK120201  
HM045801  
HM045799  
MN974224  
KX881784  
MK551553  
MK551552  
MK516540  
JN558836  
JN558835  
JN558834  
LC259093  
KX262993  
KX262989  
KX262997  
MH400249  
MG664850  
MN402885  
MK473621  
MK473622  
MK473623  
MK473625  
EU0372006  
MK120198  
MK120199  
MK120197  
MK120200  
MN974211  
MN974209  
MN974208  
MN974206  
MN974205  
MN974223  
MN974222  
MN974221  
MN974220  
MN974219  
MN974218  
MN974217  
MK473628  
MK473627  
MT526807  
MT526805  
MT526804  
MT526802  
MT526800  
MN974212  
MK473624  
MH423804  
MH423805  
MH423806  
MH423803  
MH423802  
MH423801  
MH423800  
MH423799  
MH423798  
MH423797  
MN974213  
MN974207  
MK473626  
MN402882  
MN402886  
MN974210  
MN974204  
MN402884  
MN974203  
MN402889  
JX088705  
MN402883





## Apêndice C: Localização dos códigos de barras nas proteínas modeladas.

Figura 2: CHIKV - E1 - BarCODE 1

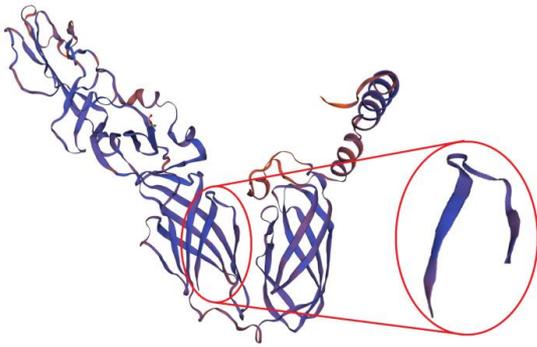


Figura 10: CHIKV - E1 - BarCODE 2

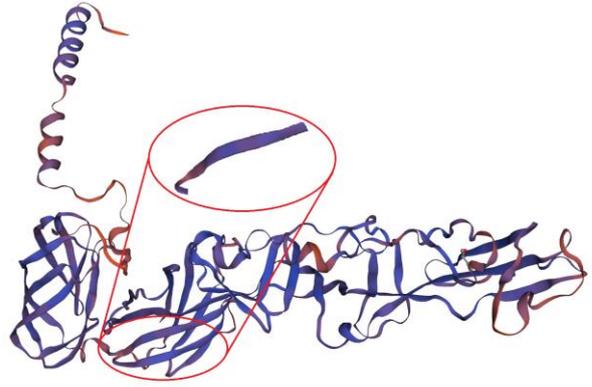


Figura 11: CHIKV - NSP2 - BarCODE 1

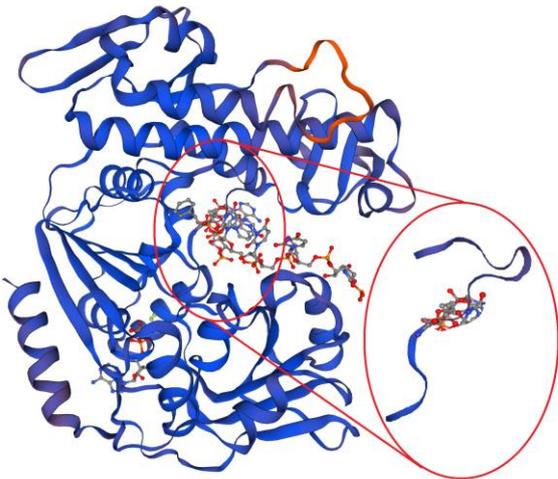


Figura 12: CHIKV - NSP2 - BarCODE 2

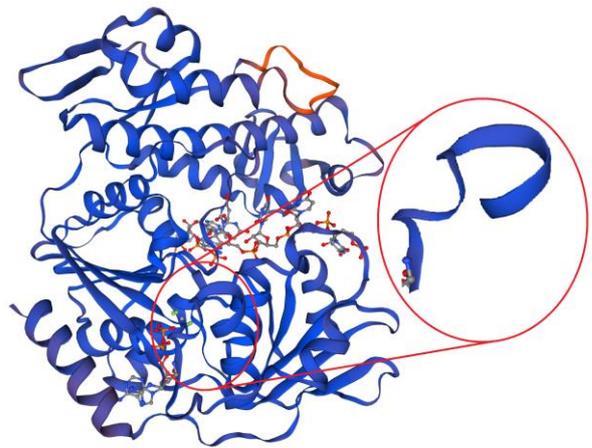


Figura 3: DENV1 - NSP3 - BarCODE 1

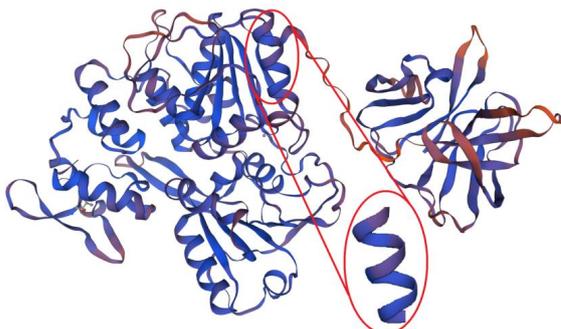


Figura 13: DENV1 - NSP3 - BarCODE 2

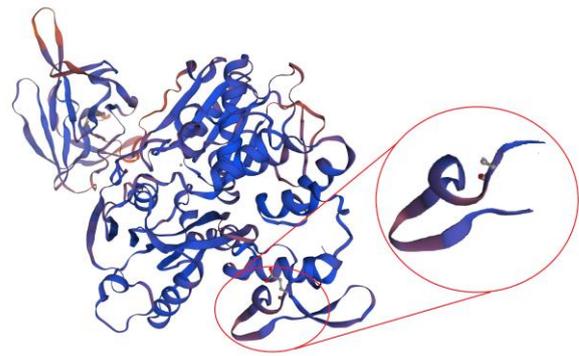


Figura 14: DENV1 - NSP5 - BarCODE 1

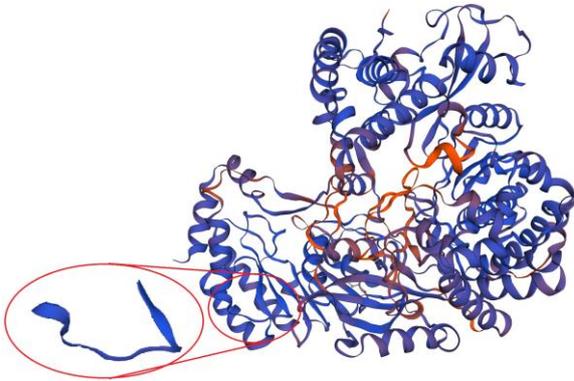


Figura 15: DENV1 - NSP5 - BarCODE 2

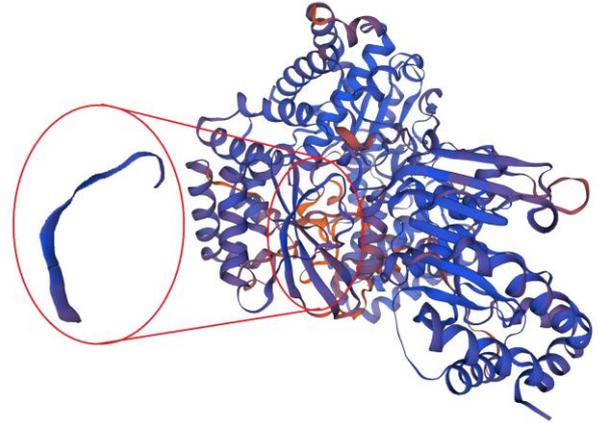


Figura 16: DENV2 - NSP3 - BarCODE 1

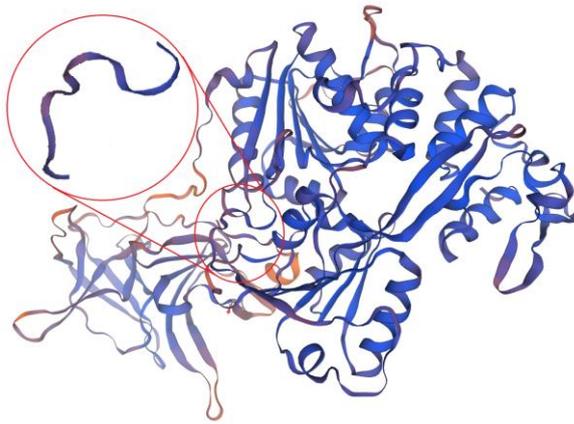


Figura 17: DENV2 - NSP3 - BarCODE 2

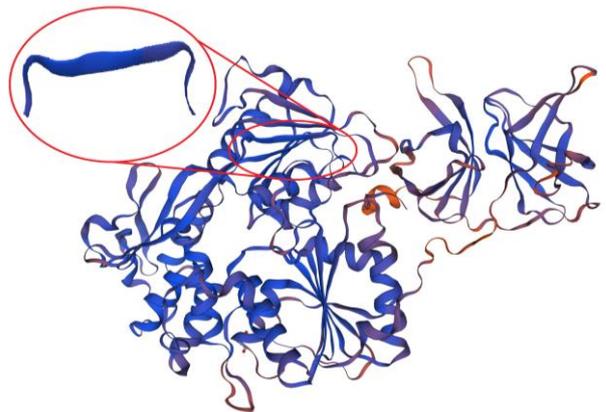


Figura 18: DENV2 - NSP5 - BarCODE 1

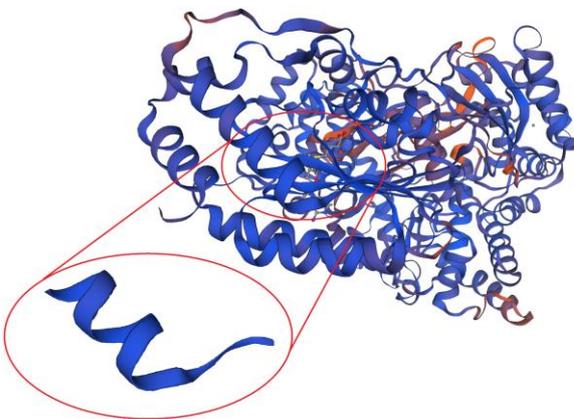


Figura 4: DENV2 - NSP5 - BarCODE 2

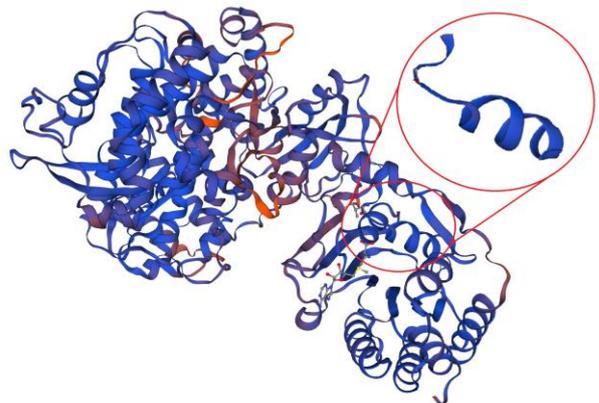


Figura 19: DENV3 - NSP3 - BarCODE 1

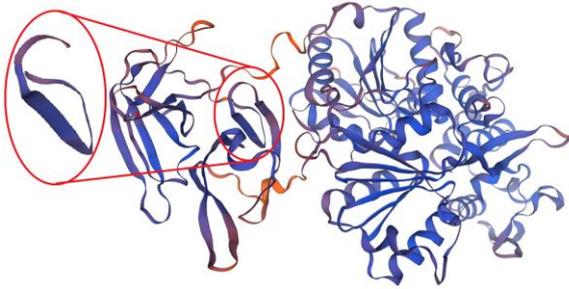


Figura 20: DENV3 - NSP3 - BarCODE 2

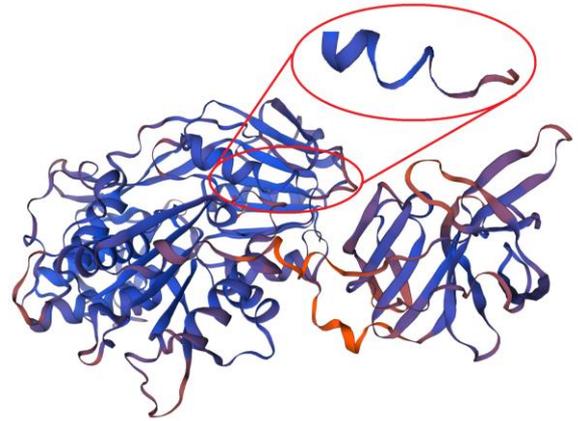


Figura 21: DENV3 - NSP5 - BarCODE 1

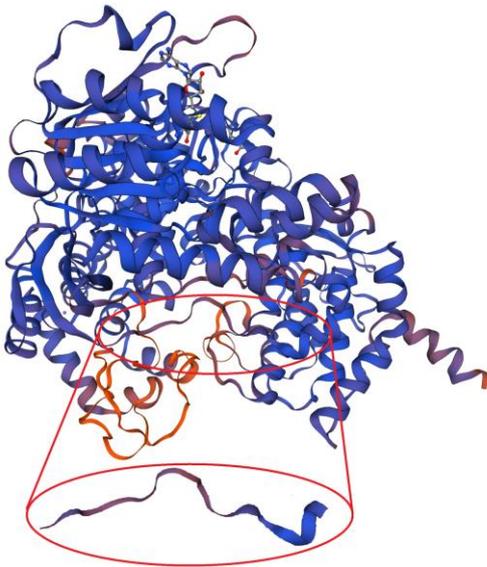


Figura 22: DENV3 - NSP5 - BarCODE 2

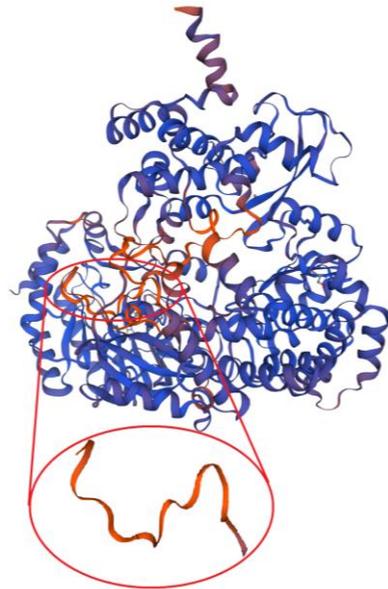
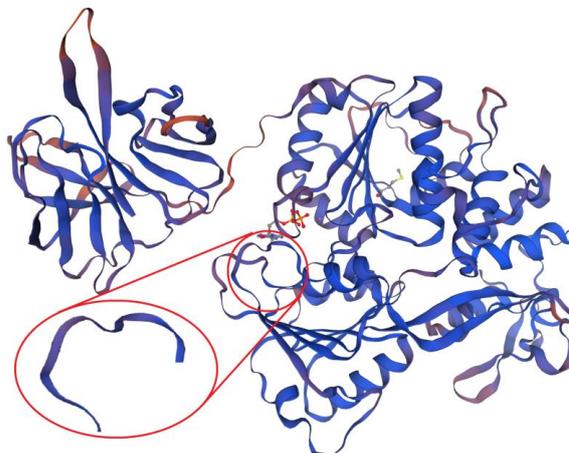


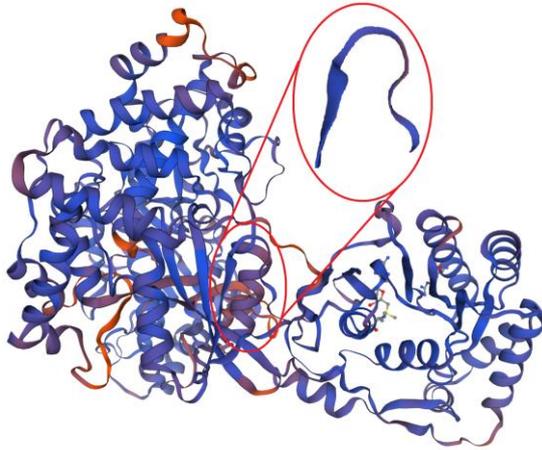
Figura 5: DENV4 - NSP3 - BarCODE 1



DENV4 - NSP3 - BarCODE 2

Não foi determinado um segundo código de barras para o DENV4 utilizando o gene NSP3.

Figura 23: DENV4 - NSP5 - BarCODE 1



DENV4 - NSP5 - BarCODE 2

Não foi determinado um segundo código de barras para o DENV4 utilizando o gene NSP5.

Figura 24: MAYV - E1 - BarCODE 1

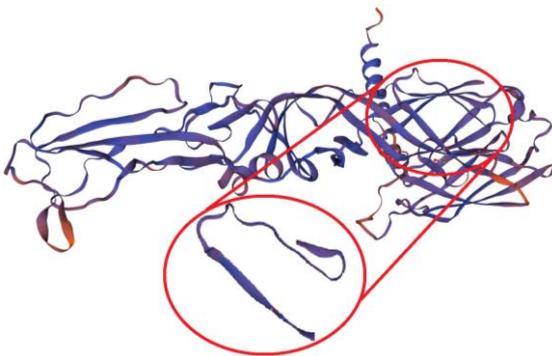


Figura 25: MAYV - E1 - CarCODE 2

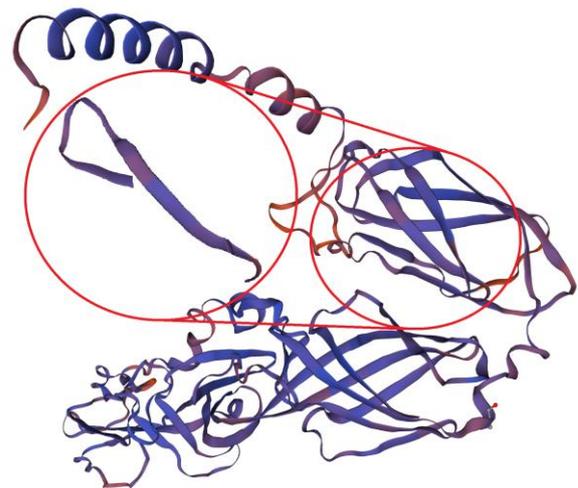


Figura 6: MAYV - NSP2 - BarCode 1

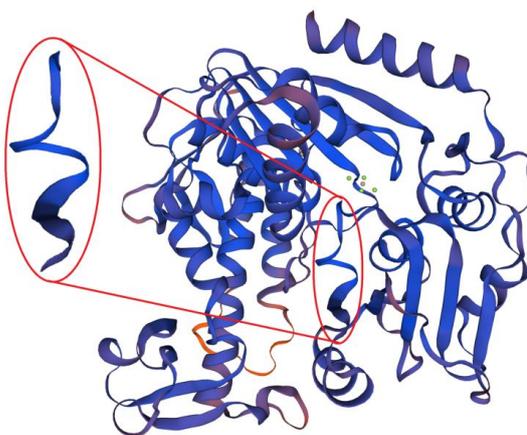


Figura 26: MAYV - NSP2 - BarCODE2

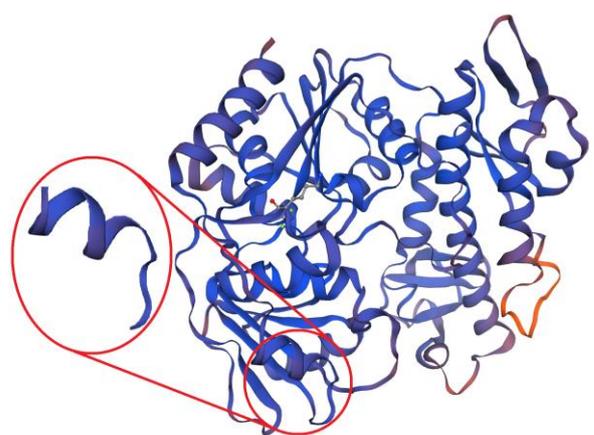


Figura 27: WNV - NSP3 - BarCODE 1

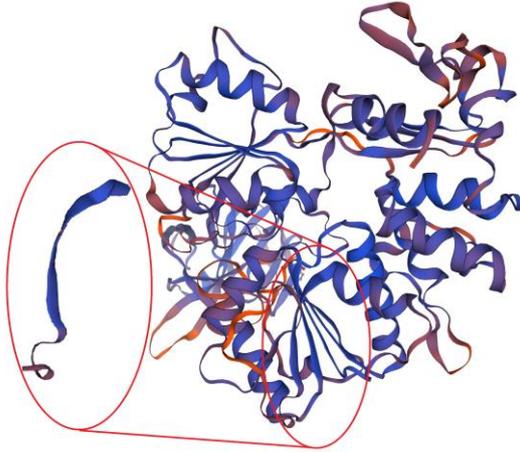


Figura 28: WNV - NSP3 - BarCODE 2

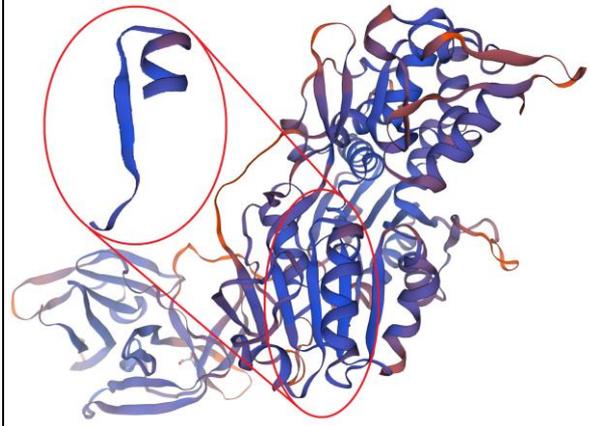


Figura 7: WNV - NSP5 - BarCODE 1

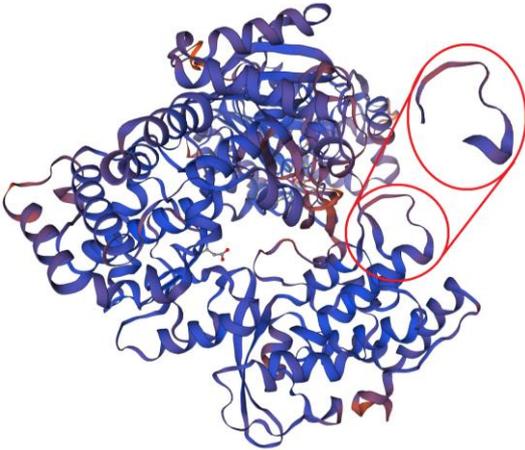


Figura 29: WNV - NSP5 - BarCODE 2

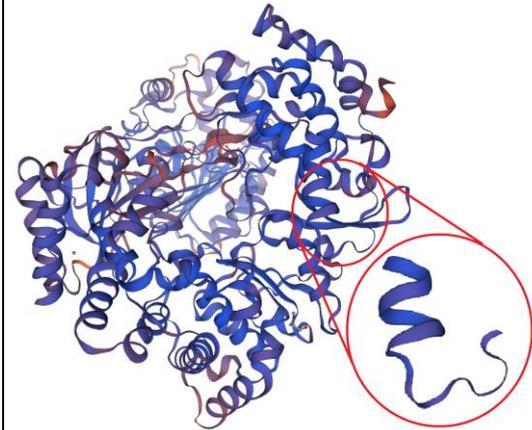


Figura 30: YFV - NSP5 - BarCODE 1

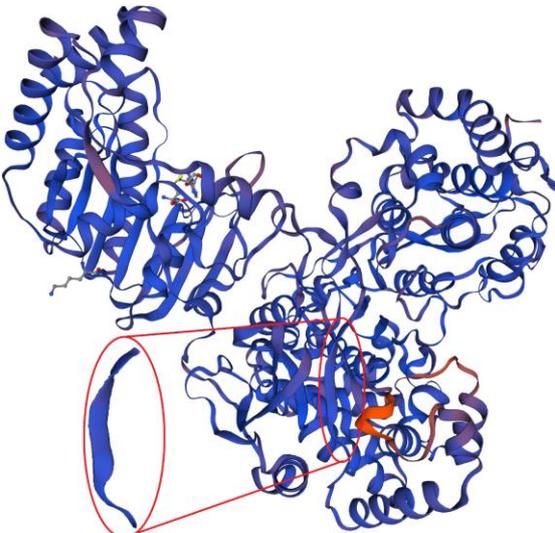
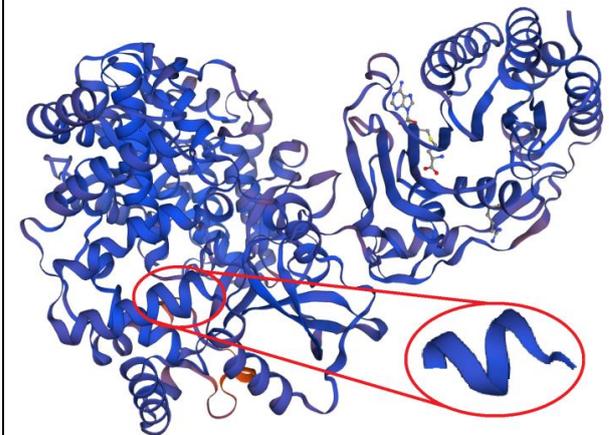


Figura 8: YFV - NSP5 - BarCODE 2



YFV - prM - BarCODE 1

Não foram determinados códigos de barras para o YFV utilizando o gene prM.

YFV - prM - BarCODE 2

Não foram determinados códigos de barras para o YFV utilizando o gene prM.

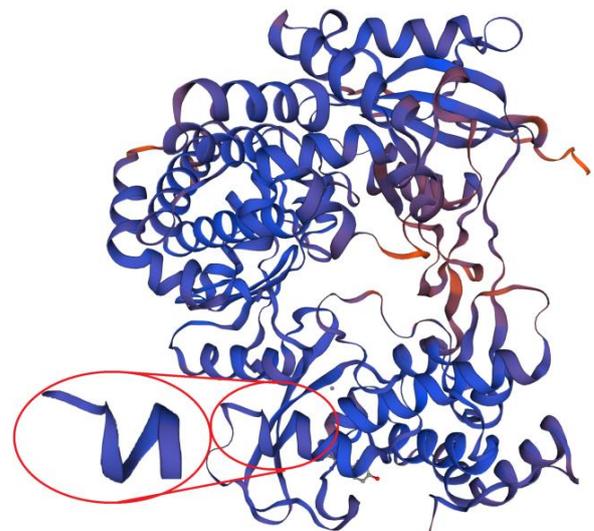
Figura 31: ZIKV - NSP5 - BarCODE 1



ZIKV - prM - BarCODE 1

Não foram determinados códigos de barras para o ZIKV utilizando o gene prM.

Figura 9: ZIKV - NSP5 - BarCODE 2



ZIKV - prM - BarCODE 2

Não foram determinados códigos de barras para o ZIKV utilizando o gene prM.