

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Escola de Veterinária
Programa de Pós-Graduação em Ciência Animal

Bruna Ferreira Sampaio Ribeiro

IDENTIFICAÇÃO DE NOVOS ALVOS PARA DIAGNÓSTICO DE
***SENECAVIRUS A* E PRODUÇÃO DE VACINAS**
A PARTIR DE ANÁLISES *IN SILICO*

Belo Horizonte

2024

Bruna Ferreira Sampaio Ribeiro

**IDENTIFICAÇÃO DE NOVOS ALVOS PARA DIAGNÓSTICO DE
SENECAVIRUS A E PRODUÇÃO DE VACINAS
A PARTIR DE ANÁLISES *IN SILICO***

Dissertação apresentada ao Programa de Pós-Graduação em Ciência Animal da Escola de Veterinária da Universidade Federal de Minas Gerais, como requisito para obtenção do título de Mestre em Ciência Animal.

Orientadora: Prof^ª. Dr^ª. Maria Isabel
Maldonado Coelho Guedes

Co-orientadora: Prof^ª. Dr^ª. Flávia Figueira
Aburjaile

Belo Horizonte

2024

R484i Ribeiro, Bruna Ferreira Sampaio, 1994 -
Identificação de novos alvos para diagnóstico de Senecavirus A e
produção de vacinas a partir de análises in Silico/ Bruna Ferreira Sampaio
Ribeiro.-2024.
131 f. il.

Orientadora: Maria Isabel Maldonado Coelho Guedes
Coorientadora: Flávia Figueira Aburjaile

Dissertação (Mestrado) apresentada à Faculdade de Medicina Veterinária da
UFMG, como requisito parcial para obtenção do título de Mestra em Ciência
Animal.

Bibliografia: f. 105 - 118.

1. Suíno - Doenças - Teses - 2. Febre aftosa - Teses - I. Guedes,
Maria Isabel Maldonado Coelho - II. Aburjaile, Flávia Figueira -
III. Universidade Federal de Minas Gerais, Escola de Veterinária
IV. Título.

CDD – 636.089 26

Bibliotecária responsável Cristiane Patrícia Gomes CRB 2569
Biblioteca da Escola de Veterinária, UFMG.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE VETERINÁRIA
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL

FOLHA DE APROVAÇÃO

BRUNA FERREIRA SAMPAIO RIBEIRO

Dissertação submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA ANIMAL, como requisito para obtenção do grau de MESTRE em CIÊNCIA ANIMAL, área de concentração **MEDICINA VETERINÁRIA PREVENTIVA**.

Aprovado(a) em 16 de fevereiro de 2024, pela banca constituída pelos membros:

Dr.(a). Maria Isabel Maldonado Coelho Guedes - Orientador(a)

Dr.(a). Erica Azevedo Costa

Dr.(a). Marcelo Fernandes Camargos



Documento assinado eletronicamente por **Maria Isabel Maldonado Coelho Guedes, Professora do Magistério Superior**, em 22/02/2024, às 08:33, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Erica Azevedo Costa, Professora do Magistério Superior**, em 22/02/2024, às 08:44, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marcelo Fernandes Camargos, Usuário Externo**, em 23/02/2024, às 09:19, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3024837** e o código CRC **690E9F48**.

AGRADECIMENTOS

Aos meus pais, pelo amor incondicional e por todos os esforços realizados para meu conforto e felicidade. Ao meu marido, Pedro Henrique, por todo o carinho, pelo apoio diário e por contribuir em todos os sentidos para o meu crescimento. Ao Di, Gina, Paulo e Isabela, pelo entusiasmo diante de cada conquista. À professora Maria Isabel, pela disponibilidade de orientar este trabalho e por acolher e incentivar minhas idealizações. À professora Flávia, pelos ensinamentos e oportunidades. À professora Érica, por toda a leveza e consideração que transmite. Aos amigos do LPVA, pela parceria, companheirismo e afeto. Aos alunos do LBI, por todo o auxílio durante o desenvolvimento do projeto. Aos amigos que se fazem presentes mesmo diante das inevitáveis ausências. À Nanda, companheira fiel nos dias bons e ruins. E, claro, à Suri, que carinhosamente me acompanhou nas alegrias e desafios dos períodos mais importantes da minha vida, cumpriu sua missão com ternura e se despediu.

RESUMO

A infecção por *Senecavirus A* (SVA) é uma das doenças vesiculares clinicamente indistinguíveis de febre aftosa, sendo comumente detectada em estabelecimentos de abate de suínos a partir de lesões em cicatrização ou cicatrizadas. Desta forma, implica em prejuízos tanto diretos, por queda na produtividade decorrente das vesículas e por mortes neonatais; quanto prejuízos indiretos, diante das restrições comerciais e do dispêndio de recursos para investigação pelo serviço veterinário oficial, resultantes da similaridade clínica com a febre aftosa. O Ministério da Agricultura e Pecuária (MAPA) determina as medidas a serem adotadas diante de suspeita de doença vesicular, como a coleta de material para diagnóstico, a investigação epidemiológica que busque determinar a origem e o risco de disseminação do evento sanitário; e a adoção de medidas de biossegurança. Ressalta-se que o Laboratório de Defesa Agropecuária (LFDA), situado em Pedro Leopoldo – MG, é o único laboratório de nível de biossegurança 4 reconhecido pela World Organisation for Animal Health (WOAH) no Brasil, estando exclusivamente autorizado a realizar o diagnóstico de doenças vesiculares no país. Além de dificuldades referentes às longas distâncias entre algumas regiões e o acesso a este laboratório, os testes sorológicos e moleculares preconizados no Programa Nacional de Vigilância para a Febre Aftosa (PNEFA) são laboriosos e demandam equipamentos específicos e pessoal treinado, gerando morosidade ao processo investigativo. Neste contexto, a prevenção da circulação de *Senecavirus A* e do aparecimento de sintomas por meio da vacinação, evitando-se a necessidade de investigações, bem como a eficácia diagnóstica para diferenciar de forma célere a infecção por SVA da infecção pelo vírus da febre aftosa, são pontos fundamentais para a prosperidade da suinocultura nacional e internacional. Assim, o objetivo do trabalho foi encontrar moléculas-alvo adequadas para o diagnóstico sensível e específico da infecção por *Senecavirus A*, e possíveis antígenos para formulação vacinal. Ferramentas de bioinformática foram utilizadas para realizar análise filogenética de sequências de SVA disponíveis nos bancos de dados e para prever epítomos de linfócitos B e ligantes de MHC de classes I e II. Os epítomos identificados foram utilizados para construção de duas proteínas multi-epítomo, com bons resultados nas análises de antigenicidade, alergenicidade, toxicidade, estrutura e parâmetros físico-químicos. Assim, as proteínas construídas possuem potencial para uso em testes diagnósticos e formulações vacinais, o qual deverá ser confirmado futuramente por análises *in vitro* e *in vivo*.

Palavras-chave: *Senecavirus A*; proteína multi-epítomo; predição *in silico*; diagnóstico sorológico; vacina multi-epítomo.

ABSTRACT

Senecavirus A (SVA) infection is one of the vesicular diseases clinically indistinguishable from foot and mouth disease, and is commonly detected in swine slaughter establishments from healing or healed lesions. In this way, it implies both direct losses, due to a drop in productivity because of vesicles and neonatal deaths; as well as indirect losses, due to commercial restrictions and the expenditure of resources for investigation by the official veterinary service, because of the clinical similarity with foot-and-mouth disease. The Ministry of Agriculture and Livestock (MAPA) determines the measures to be adopted in the event of suspect vesicular disease, such as the collection of material for diagnosis, epidemiological investigation to determine the origin and risk of dissemination of the health event; and the adoption of biosecurity measures. It is noteworthy that the Agricultural Defense Laboratory (LFDA), located in Pedro Leopoldo – MG, is the only biosafety level 4 laboratory recognized by the World Organisation for Animal Health (WOAH) in Brazil, being exclusively authorized to carry out the diagnosis of vesicular diseases in the country. In addition to difficulties related to the long distances between some regions and access to this laboratory, the serological and molecular tests recommended in the National Foot and Mouth Disease Surveillance Program (PNEFA) are laborious and require specific equipment and trained personnel, causing slowness in the investigative process. In this context, the prevention of the circulation of *Senecavirus A* and the appearance of symptoms through vaccination, avoiding the need for investigations, as well as the diagnostic efficacy to quickly differentiate SVA infection from foot-and-mouth disease virus infection, are fundamental points for the prosperity of national and international pig farming. Thus, the objective of the work was to find suitable target molecules for the sensitive and specific diagnosis of *Senecavirus A* infection, and possible antigens for vaccine formulation. Bioinformatics tools were used to perform phylogenetic analysis of SVA sequences available in databases and to predict B lymphocyte epitopes and MHC class I and II ligands. The identified epitopes were used to construct two multi-epitope proteins, with good results in the analyses of antigenicity, allergenicity, toxicity, structure and physicochemical parameters. Thus, the constructed proteins have potential for use in diagnostic tests and vaccine formulations, which should be confirmed in the future by *in vitro* and *in vivo* analyses.

Keywords: *Senecavirus A*; multi-epitope protein; *in silico* prediction; serological diagnosis; multi-epitope vaccine.

LISTA DE FIGURAS

Figura 1 – Representação esquemática do genoma de <i>Senecavirus A</i> (SVA) e do processamento da poliproteína.....	16
Figura 2 – Representação esquemática do SVA	18
Figura 3 - Lesões apresentadas por porcas naturalmente infectadas por SVA.	23
Figura 4 - Lesões identificadas em leitões infectados com SVA.....	24
Figura 5 - Dinâmica da infecção por SVA.....	26
Figura 6 - Representação esquemática da tecnologia <i>Illumina</i>	34
Figura 7 - Estatísticas básicas apresentadas em resultado hipotético de sequenciamento <i>Illumina</i>	37
Figura 8 - Qualidade de sequência por base, em bom resultado de sequenciamento <i>Illumina</i> ,	38
Figura 9 - Qualidade de sequência por bloco, em bom resultado <i>Illumina</i>	39
Figura 10 - <i>Score</i> de qualidade por sequência, em resultado de sequenciamento <i>Illumina</i>	40
Figura 11 - Conteúdo de sequência por base, em resultado de sequenciamento <i>Illumina</i>	41
Figura 12 - Conteúdo GC por base, em resultado de sequenciamento <i>Illumina</i>	42
Figura 13 - Conteúdo N por base, em resultado de sequenciamento <i>Illumina</i>	43
Figura 14 - Distribuição do comprimento das sequências, em resultado de sequenciamento <i>Illumina</i>	44
Figura 15 - Nível de sequências duplicadas, em resultado de sequenciamento <i>Illumina</i>	45
Figura 16 - Sequências super-representadas.	46
Figura 17 - Conteúdo de adaptadores.....	47
Figura 18 - Representação esquemática das estruturas de uma árvore filogenética.	50
Figura 19 - Representação das estruturas secundárias de proteínas por quadrante do <i>Ramachandran plot</i>	60
Figura 20 - Estatísticas básicas do sequenciamento da amostra LPVA4.....	67
Figura 21 - Qualidade da sequência LPVA4 por base	68
Figura 22 - Qualidade da sequência LPVA4 por bloco	69
Figura 23 - <i>Scores</i> de qualidade da sequência LPVA4.	70
Figura 24 - Conteúdo da sequência LPVA4por base	71
Figura 25 - Conteúdo GC da sequência LPVA4.....	72
Figura 26 - Conteúdo N d sequência LPVA4	73
Figura 27 - Distribuição do comprimento da LPVA4.....	74
Figura 28 - Nível de duplicação da LPVA4.....	75
Figura 29 - Sequências super-representadas da LPVA4	76
Figura 30 - Conteúdo de adaptadores da amostra LPVA4.....	77
Figura 31 - Conteúdo K-mer da sequência LPVA4	78
Figura 32 - Dados das sequências K-mer da LPVA4	79
Figura 33 - Sequência da amostra LPVA4, obtida pelo software BWA.....	81
Figura 34 - Predição do melhor modelo evolutivo.....	82
Figura 35 - Árvore de máxima verossimilhança construída a partir do modelo GTR + I + G, com 1000 <i>bootstraps</i>	83
Figura 36 – Distância genética entre as sequências brasileiras disponíveis no NCBI <i>Virus</i>	84
Figura 37 - Sequências completas de SVA depositadas ao longo dos anos.....	85
Figura 38 – Sequências completas de SVA depositadas por país.....	86
Figura 39 - Alinhamento das sequências de proteínas de SVA oriundas do Brasil.....	89

Figura 40 - Análise de homologia entre o proteoma de suínos e as proteínas preditas ..	93
Figura 41 – Análise de homologia entre a proteína M-E1 e outros vírus causadores de doença vesicular.	94
Figura 42 - Análise de homologia entre a proteína M-E2 e outros vírus causadores de doença vesicular	95
Figura 43 - Estrutura secundária da proteína M-E1	96
Figura 44 - Estrutura terciária inicialmente predita para a proteína M-E1.	96
Figura 45 - <i>Ramachandran plot</i> da estrutura inicialmente predita para a proteína M-E1	97
Figura 46 - Estrutura terciária da proteína M-E1, após refinamento	98
Figura 47 - <i>Ramachandran plot</i> da estrutura refinada da proteína M-E1	99
Figura 48 - Estrutura secundária da proteína M-E2	100
Figura 49 - Estrutura terciária inicialmente predita para a proteína M-E2.	100
Figura 50 - <i>Ramachandran plot</i> da estrutura inicialmente predita para a proteína M-E2	101
Figura 51 - Estrutura terciária da proteína M-E2, após refinamento	102
Figura 52 - <i>Ramachandran plot</i> da estrutura refinada da proteína M-E2	102

LISTA DE TABELAS

Tabela 1 - Vantagens e desvantagens das principais plataformas de sequenciamento ...	36
Tabela 2 - Principais características dos métodos de construção de árvores filogenéticas	55
Tabela 3 - Dados dos epítomos de linfócito B selecionados.....	88
Tabela 4 - Resultados das análises da proteína multi-epítopo M-E1, com e sem adjuvantes.	90
Tabela 5 - Resultados das análises da proteína multi-epítopo M-E2, construída com diferentes opções de <i>linkers</i>	91
Tabela 6 - Parâmetros físico-químicos das proteínas multi-epítopo M-E1 e M-E2.	92

LISTA DE APÊNDICES

Apêndice I – Dados das sequências de nucleotídeos de SVA disponíveis no NCBI <i>Virus</i> utilizadas no trabalho	119
Apêndice II – Dados das sequências de proteína de SVA disponíveis no NCBI <i>Virus</i> e utilizadas no trabalho.	124
Apêndice III – Dados das sequências brasileiras de SVA disponíveis no NCBI <i>Virus</i> utilizadas no trabalho.....	130

SUMÁRIO

1	INTRODUÇÃO	13
2	OBJETIVOS	15
2.1	Objetivo geral	15
2.2	Objetivos específicos	15
3	REVISÃO DE LITERATURA	15
3.1	<i>Senecavirus A</i>	15
3.1.1	Taxonomia e principais características	15
3.1.2	Caracterização funcional das proteínas	17
3.1.3	Ciclo de replicação	20
3.1.4	Epidemiologia	21
3.1.5	Patogenia	23
3.1.6	Imunidade	25
3.1.7	Diagnóstico	27
3.1.8	Prevenção e controle	30
3.2	Ferramentas de bioinformática	32
3.2.1	Métodos de sequenciamento genômico	32
3.2.1.1	Sequenciamento de amostras de <i>Senecavirus A</i>	36
3.2.2	Processamento dos dados do sequenciamento	37
3.2.3	Análises de filogenia	49
3.2.3.1	Análises de filogenia para <i>Senecavirus A</i>	55
3.2.4	Predições <i>in silico</i>	57
3.2.4.1	Predições <i>in silico</i> para <i>Senecavirus A</i>	61
4	MATERIAL E MÉTODOS	62
4.1	Amostra viral	62
4.2	Preparo da amostra para sequenciamento	63
4.3	Sequenciamento da amostra e análise de qualidade	63
4.4	Mapeamento e montagem do genoma	63
4.5	Alinhamento e análise filogenética	64
4.6	Predição de epítomos e avaliação da conservação	64
4.7	Construção de proteínas multi-epítopo	65
4.8	Análises de antigenicidade, alergenicidade e toxicidade	65
4.9	Avaliação de parâmetros físico-químicos	65
4.10	Análises de homologia	66
4.11	Análises estruturais	66
4.12	Predição de epítomos conformacionais de células B	66
5	RESULTADOS E DISCUSSÃO	66
5.1	Preparo da amostra para sequenciamento	66
5.2	Sequenciamento da amostra e análise de qualidade	67
5.3	Montagem do genoma	80

5.4	Alinhamento e análise filogenética.....	82
5.5	Predição de epítomos e avaliação da conservação	86
5.6	Construção de proteínas multi-epítopo.....	90
5.7	Análises de antigenicidade, alergenicidade e toxicidade.....	90
5.8	Avaliação de parâmetros físico-químicos e homologia com o hospedeiro ..	91
5.9	Análises de homologia.....	92
5.10	Análises estruturais.....	96
5.11	Predição de epítomos conformacionais de células B.....	103
6	CONCLUSÃO.....	103
7	PERSPECTIVAS.....	103
8	REFERÊNCIAS.....	105
	APÊNDICE I - Dados das 279 sequências de SVA disponíveis no NCBI <i>Virus</i> , com tamanho superior a 7000pb, utilizadas no presente trabalho (acesso em 30/10/2023).	119
	APÊNDICE II - Dados das 316 sequências de proteína de SVA disponíveis no NCBI <i>Virus</i> , com tamanho superior a 2000 aminoácidos, utilizadas no presente trabalho (acesso em 10/01/2024).	124
	APÊNDICE III - Dados das sequências brasileiras de SVA disponíveis no NCBI <i>Virus</i> , com tamanho superior a 7000 pares de base e 2000 aminoácidos, utilizadas no presente trabalho (acesso em 30/10/2023 e 10/01/2024, respectivamente).	130

1 INTRODUÇÃO

O *Senecavirus A* (SVA) foi incidentalmente identificado em 2002, no estado de Maryland, nos Estados Unidos (EUA), como contaminante da linhagem celular PER.C6 (originária de células retiniais embrionárias humanas). A identificação ocorreu devido ao efeito citopático atipicamente rápido observado nestas células, que levou a análises por SDS-PAGE, sequência N-terminal, microscopia eletrônica e estudos ultraestruturais, até que o agente citopático foi clonado em placa e inicialmente denominado *Seneca Valley virus*, em alusão ao parque *Seneca Creek State Park*, próximo ao laboratório onde ocorreu o isolamento. O sequenciamento completo do vírus foi posteriormente realizado e seu genoma foi depositado nos bancos de dados sob a denominação SVV-001 (acesso *Genbank*: [NC_013311](#)) (Hales *et al.*, 2008; Leme *et al.*, 2017).

A correlação do vírus com casos de doença vesicular em suínos, entretanto, só ocorreu a partir de 2007, com a identificação do RNA de *Senecavirus A* em amostras biológicas coletadas de animais sintomáticos durante o transporte do Canadá para os Estados Unidos (Pasma *et al.*, 2008). Em 2012, um novo caso de lesão vesicular por SVA foi detectado em um animal no estado de Indiana (EUA). Os primeiros surtos por doença vesicular e mortalidade neonatal associados ao SVA foram relatados a partir de 2014, ano em que o agente foi identificado no Brasil (Vannucci *et al.*, 2015). Atualmente, a doença é considerada endêmica em algumas regiões do país, sendo comumente detectada nos estabelecimentos de abate a partir de lesões em cicatrização ou recentemente cicatrizadas (Brasil, 2020).

A importância do vírus pode ser compreendida a partir dos prejuízos diretos e indiretos causados. De forma direta, a infecção provoca perda de desempenho dos animais, diante da dor provocada pela presença das vesículas, localizadas principalmente nos cascos, focinho e lábios; além de casos de mortalidade neonatal, que atingiram até 70% desta categoria nos primeiros surtos brasileiros por *Senecavirus A* (Vannucci *et al.*, 2015). Indiretamente, o SVA tem sua importância sustentada pela similaridade com as lesões causadas pelo vírus da febre aftosa, doença de notificação obrigatória que exige medidas severas para sua contenção.

Diante de suspeita de doença vesicular, são impostas pelo Ministério da Agricultura e Pecuária (MAPA) a coleta de material para diagnóstico, a investigação para determinar o início, a origem e o risco de disseminação do evento sanitário; e a adoção de medidas de biossegurança, como interdição da propriedade e proibição da saída de animais e de produtos que ofereçam risco para disseminação da doença (Brasil, 2020). Neste contexto, qualquer indício de doença vesicular leva à paralisação das atividades da granja e/ou frigorífico até que se descarte a

hipótese de febre aftosa, causando perdas econômicas diretas e o estremecimento das relações comerciais com países importadores da carne brasileira.

O impacto financeiro poder ser compreendido pela posição de destaque do Brasil na suinocultura mundial, considerado o quarto maior produtor e exportador de carne suína no mundo. Em 2022, a produção foi de 4,983 milhões de toneladas, com exportação de 1,120 milhão de toneladas para 88 países, equivalente a US\$2,5 bilhões (ABPA, 2023).

A indistinguibilidade clínica das lesões causadas pelo vírus da febre aftosa, SVA e outras doenças como estomatite vesicular, doença vesicular dos suínos e exantema vesicular dos suínos (as duas últimas consideradas exóticas no Brasil), impõe a utilização de métodos diagnósticos para confirmação do patógeno associado ao quadro. Ressalta-se que o Laboratório Federal de Defesa Agropecuária (LFDA), situado em Pedro Leopoldo – MG, está exclusivamente autorizado a realizar o diagnóstico de doenças vesiculares no país, por atender ao nível de biossegurança 4 dentro dos padrões reconhecidos pela *World Organisation for Animal Health* (WOAH) (Brasil, 2017).

A realização do diagnóstico de doença vesicular por um único laboratório implica em morosidade ao processo investigativo e exposição prolongada aos prejuízos já descritos, não apenas pela distância entre algumas regiões e o LFDA, o que pode prejudicar a adequada conservação das amostras e a eficácia diagnóstica, mas também pelas características inerentes aos testes de soroneutralização e de reação de transcriptase reversa seguida de reação em cadeia da polimerase (convencional – RT-PCR; ou em tempo real – RT-qPCR), preconizados no Programa Nacional de Vigilância para a Febre Aftosa (PNEFA) (Brasil, 2020; Brasil, 2022), que demandam tempo, equipamentos e treinamentos específicos.

O cenário atual de retirada progressiva da vacinação contra febre aftosa no país reforça a importância da investigação do agente etiológico envolvido no quadro de doença vesicular e do fortalecimento do sistema de vigilância. Isso porque falhas ou lentidão na investigação colocariam em risco o status de país livre de febre aftosa sem vacinação, pretendido pelo Plano Estratégico 2017-2026 do PNEFA, e causariam prejuízos imensuráveis ao agronegócio brasileiro, especialmente considerando-se os receios existentes quanto ao alto poder de difusão do vírus e à maior susceptibilidade de animais diante da suspensão da vacinação (Brasil, 2017).

Neste contexto, a prevenção da circulação de *Senecavirus A* e do aparecimento de sintomas por meio da vacinação, evitando-se a necessidade de investigações; bem como a eficácia diagnóstica para diferenciar de forma célere a infecção por SVA da infecção pelo vírus da febre aftosa, são pontos fundamentais para a prosperidade da suinocultura nacional e

internacional. Diversas pesquisas vêm sendo realizadas neste sentido, e a bioinformática se apresenta como uma ferramenta viável para a identificação de alvos vacinais e de diagnóstico.

2 OBJETIVOS

2.1 Objetivo geral

Identificar novos alvos para a detecção sensível e específica de *Senecavirus A*, que posteriormente possam ser utilizados para o desenvolvimento de testes diagnósticos, bem como potenciais proteínas para desenvolvimento de vacina.

2.2 Objetivos específicos

- Sequenciar e mapear o genoma da amostra de SVA disponível no Laboratório de Pesquisa em Virologia Animal (LPVA);
- Realizar alinhamento e análise filogenética com o genoma sequenciado e as principais sequências de nucleotídeos disponíveis no NCBI (*National Center for Biotechnology Information*);
- Selecionar epítomos-alvo de células B de regiões imunogênicas conservadas do genoma de *Senecavirus A* para posterior desenvolvimento de testes sorológicos;
- Selecionar ligantes do complexo principal de histocompatibilidade (MHC) de classe I e II para construção de proteína multi-epítomo a ser usada como antígeno vacinal.

3 REVISÃO DE LITERATURA

3.1 *Senecavirus A*

3.1.1 Taxonomia e principais características

O sequenciamento do genoma do SVV-001 permitiu sua caracterização como um vírus típico da família *Picornaviridae*, com o padrão L-4-3-4 (*Leader* – 4 polipeptídeos da proteína P1 – 3 polipeptídeos da proteína P2 – e 4 polipeptídeos da proteína P3), apresentado na figura 1. Apesar de compartilhar características principalmente com o gênero *Cardiovirus*, no que diz respeito aos polipeptídeos P1, 2C, 3C e 3D, existem diferenças relacionadas aos polipeptídeos 2B e 3A e ao sítio interno de entrada dos ribossomos (*internal ribosomal entry site* – IRES), que é do tipo IV no *Senecavirus A* e do tipo II nos *Cardiovirus*. Além disso, diferenças significativas nas regiões genômicas 2A, 2B, 3A e 3B de outros picornavírus justificaram que o SVA fosse considerado uma nova espécie dentro de um novo gênero (Hales *et al.*, 2008; Leme *et al.*, 2017).

Esta classificação foi confirmada pelo *International Committee on Taxonomy of Viruses* (ICTV) em 2009, ocasião na qual foi criado o gênero *Senecavirus* e a espécie *Seneca Valley virus*. A classificação mais recente dada pelo ICTV, em 2022, reconhece o SVA como pertencente à família *Picornaviridae*, subfamília *Caphthovirinae*, gênero *Senecavirus* e espécie *Senecavirus A* - a única deste gênero (ICTV, 2022).

Como principais características, o vírus possui genoma RNA linear, de fita simples, não segmentado e de polaridade positiva, com aproximadamente 7300 nucleotídeos (nt). Apresenta formato icosaédrico e ausência de envelope (Wang *et al.*, 2023).

Conforme ilustrado na figura 1, o RNA genômico possui uma extremidade 5'UTR de aproximadamente 660 nt, que contém uma sequência IRES do tipo IV e está covalentemente ligada à proteína viral VPg; uma ORF única; uma extremidade 3' UTR poliadenilada de aproximadamente 70 nt; e uma cauda poli (A). A ORF origina uma extensa poliproteína, com cerca de 2180 aminoácidos, que consiste na proteína *Leader* (L) e em três regiões proteicas principais (P1, P2 e P3) (Liu *et al.*, 2020; Wang *et al.*, 2023).

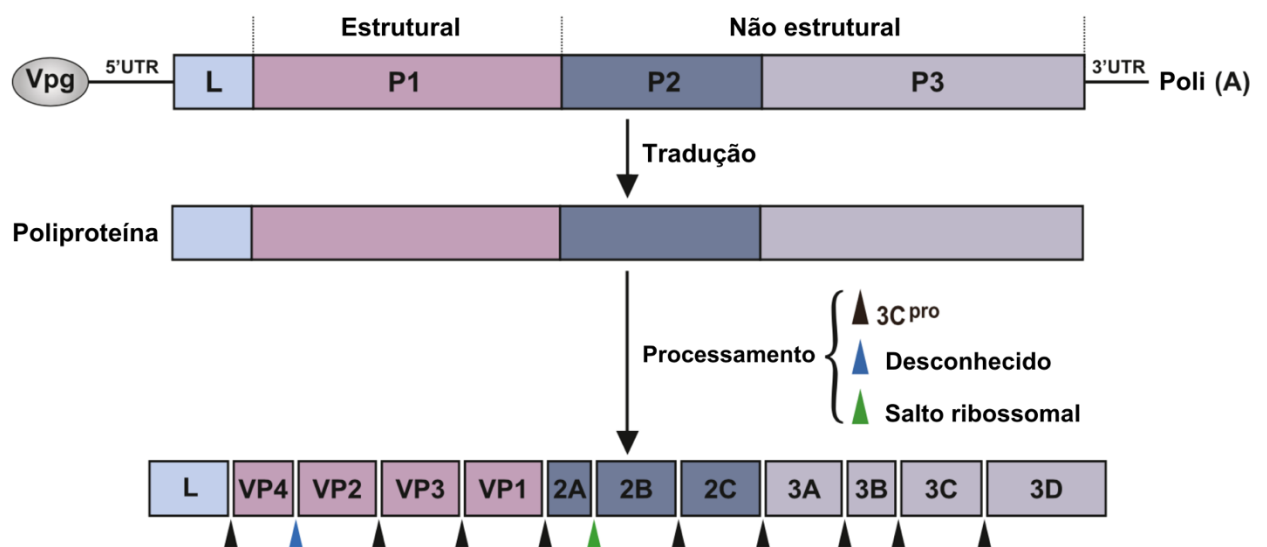


Figura 1 – Adaptada de Wang *et al.*, 2023. Representação esquemática do genoma de SVA e do processamento da poliproteína. O arranjo genômico inicia-se pela proteína Vpg, ligada covalentemente à extremidade 5' flanqueada por região não traduzida (*untranslated region* – UTR), seguida por uma única e longa sequência aberta de leitura (*open reading frame* – ORF), a extremidade 3'UTR e a cauda poli (A). A tradução da ORF resulta em uma poliproteína que pode ser processada em proteínas estruturais (VP1, VP2, VP3 e VP4) e não estruturais, a partir da clivagem mediada por 3C, salto ribossomal ou outros mecanismos desconhecidos.

A poliproteína pode ser processada pela protease 3C^{pro}, por salto ribossomal e por outras formas ainda não elucidadas. O precursor P1 é clivado pela protease 3C^{pro} em VP0, VP1 e VP3, sendo VP0 posteriormente clivada em VP2 e VP4 por um mecanismo ainda desconhecido. P2 é processado nas proteínas não estruturais 2A, 2B e 2C, ocorrendo salto ribossomal na região entre 2A e 2B; e P3 é clivado em 3A, 3B, 3C e 3D, também não estruturais. VP1, VP2 e VP3 participam da formação do capsídeo viral e estão envolvidas na ligação aos receptores celulares e entrada do vírus nas células-alvo, além de possuírem epítomos em suas superfícies, que interagem com o sistema imunológico do hospedeiro e são bons alvos para diagnósticos sorológicos e vacinas. Já as proteínas não estruturais estão envolvidas na replicação do genoma viral, no metabolismo da célula hospedeira e na evasão do sistema imune (Liu *et al.*, 2020; Wang *et al.*, 2023).

3.1.2 Caracterização funcional das proteínas

Wang e colaboradores (2023) revisaram sistematicamente as funções moleculares das proteínas estruturais e não estruturais de SVA, baseado em informações disponíveis na literatura para este e outros picornavírus.

Em relação às proteínas estruturais, sabe-se que VP1, VP2 e VP3 estão presentes na parte externa do capsídeo viral (figura 2-A), estando associadas à invasão das células-alvo e à indução de resposta imune. Especificamente em relação à VP1 e VP3, estudos recentes demonstram participação destas também na ativação de vias de sinalização para indução de autofagia (Song *et al.*, 2022). A VP2, por sua vez, além de ser alvo de anticorpos do hospedeiro, possui resíduos de aminoácidos que determinam mudanças na virulência e participam da clivagem de VP0 (Strauss *et al.*, 2018).

Quanto à imunidade celular e humoral induzida pelas proteínas VP1, VP2 e VP3, estudo de Maggioli e colaboradores (2018) demonstrou que a frequência de células T, a produção de interferon gama (IFN- γ) por estas células e os títulos de IgG contra VP2 e VP3 foram superiores àqueles contra VP1. Além disso, no que diz respeito aos títulos de IgM, estes foram direcionados principalmente contra VP2 e VP3, e não foram consistentes contra VP1. Estes resultados sugerem que VP2 e VP3 podem conter um maior número de epítomos de SVA, sendo a VP2 considerada a mais imunogênica dentre as três proteínas, baseado em ensaio sorológico realizado por Dvorak e colaboradores (2017).

A proteína estrutural VP4 está localizada na parte interna do capsídeo (figura 2-B), e estudos baseados em outros picornavírus demonstram que esta exerce função especial de formação de poros na membrana endossomal, permitindo assim a entrada do RNA viral no

citoplasma da célula hospedeira para que seja iniciada a tradução e replicação (Panjwani *et al.*, 2014).

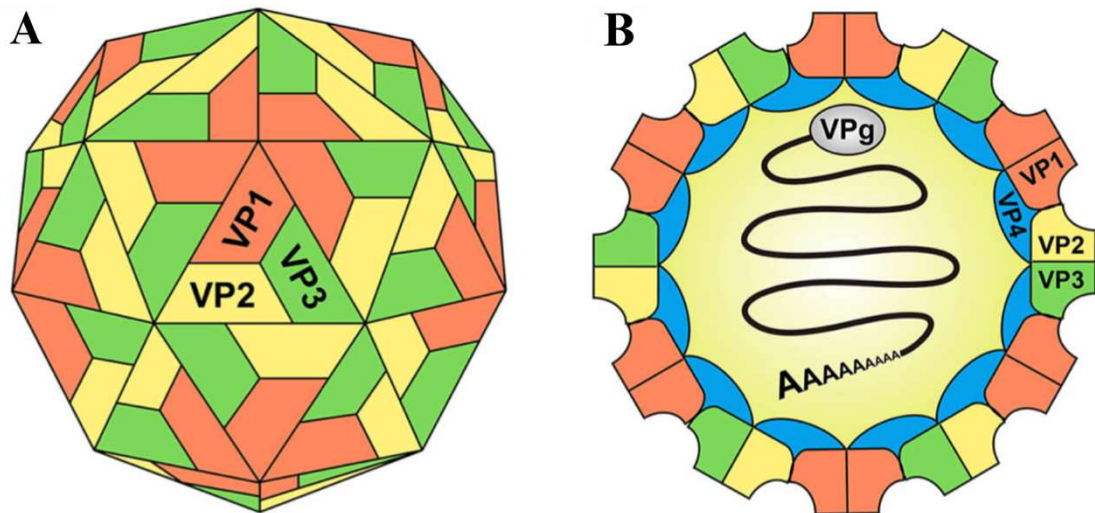


Figura 2 – Adaptada de Liu *et al.*, 2020. Representação esquemática do *Senecavirus A*. O capsídeo viral é composto por um arranjo icosaédrico de 60 protômeros, cada um formado por quatro proteínas estruturais – VP1, VP2, VP3 e VP4. As proteínas VP1, VP2 e VP3 se unem para formar o invólucro externo (A), enquanto a VP4 está localizada na face interna do capsídeo (B).

Além das proteínas estruturais, o *Senecavirus A* possui oito proteínas não estruturais, que possuem funções associadas à replicação e maturação viral, à evasão do sistema imune, e a outros aspectos ainda não compreendidos (Wang *et al.*, 2023).

A proteína *leader* (L) é a primeira codificada pelo SVA, e apresenta sequência e função variáveis dentre os diferentes picornavírus. No caso do vírus da febre aftosa, essa proteína assume função de clivagem de fatores de iniciação eucarióticos e modulação da transcrição e tradução do vírus, restringindo a síntese de proteínas da célula hospedeira. No entanto, a proteína L do SVA não possui aptidão para atividade proteolítica, e seu papel na dinâmica viral permanece indefinido (Wang *et al.*, 2023).

A presença de um *motif* NPG/P no terminal carboxila do peptídeo 2A induz um evento de co-tradução denominado “salto ribossomal” (*ribosomal skipping*), separando a proteína viral no sítio “G↓P” e, assim, promovendo a clivagem entre 2A e 2B. Até o momento não existem outras atividades conhecidas da proteína 2A (Wang *et al.*, 2023).

A proteína 2B exerce atividade semelhante à viroporina, aumentando a permeabilidade da membrana celular para promover a replicação e liberação do vírus durante a infecção celular.

Ademais, estudo recente demonstra potencial participação desta proteína na resposta imune do hospedeiro, inibindo a produção de interferon tipo I ao degradar a proteína de sinalização antiviral mitocondrial (MAVS) (Liu *et al.*, 2022a).

Estudos apresentados por Wang e colaboradores (2023) elucidam diversas funções da proteína 2C, como participação na síntese de RNA, no desnudamento viral, em rearranjos da membrana hospedeira, na replicação, na morfogênese e na evasão imunológica. Apesar de existirem evidências quanto às funções mencionadas, os mecanismos bioquímicos envolvidos em cada uma delas ainda não foram esclarecidos.

Parte do conhecimento a respeito da proteína 3A baseia-se nos estudos realizados com outros picornavírus, que sugerem o envolvimento desta proteína na replicação viral e na patogênese (Wang *et al.*, 2023). Um estudo recente, voltado para o vírus da febre aftosa (*foot-and-mouth disease virus* - FMDV), avaliou o papel da proteína 3A também para SVA, e demonstrou potencial desta em inibir a sinalização de interferon e neutralizar respostas antivirais inatas no hospedeiro por autofagia, nestes e em outros picornavírus (Yang *et al.*, 2020).

A região 3B codifica a proteína VPg, covalentemente ligada à extremidade 5' do genoma, e que atua como iniciador proteico para síntese de RNA. Além deste envolvimento na replicação, a proteína 3B também interage com fatores antivirais do hospedeiro, bloqueando a sinalização imunológica mediada por RIG-I e a resposta antiviral, tanto para FMDV quanto para SVA (Zhang *et al.*, 2020).

Diversos estudos destacados por Wang e colaboradores (2023) elucidam que as características estruturais da proteína 3C contribuem para a clivagem de poliproteínas virais e montagem do complexo de replicação viral. Além disso, esta proteína está envolvida no antagonismo da imunidade antiviral do hospedeiro, seja na degradação de moléculas envolvidas na sinalização antiviral, como RIG-I, ou na supressão de fatores reguladores de interferon (IRF3 e IRF7), dentre outros mecanismos. A morte celular também é regulada pela proteína 3C, a partir da clivagem de PARP1 e NF- κ B p65 de forma direta ou por ativação da caspase-3. A proteína 3C também atua na manipulação do mecanismo de autofagia, visando favorecer a replicação do vírus; e na inibição da maquinaria de tradução do hospedeiro, priorizando a tradução do RNA viral. Quanto à replicação viral, a proteína 3C atua por diferentes processos, como na prevenção da formação de grânulos de estresse que regulam a tradução e possuem efeito antiviral; e na degradação de fatores antivirais importantes como DDX21 e DHX30.

Por fim, a proteína 3D é uma RNA polimerase dependente de RNA (RdRp), fundamental para replicação viral e evolução genética do vírus. Esta interage com outras

proteínas virais e com fatores do hospedeiro para iniciar e regular a síntese de RNA viral, além de induzir a produção de IL-1 β , citocina pró-inflamatória associada à modulação da resposta imune inata contra infecções virais (Choudhury *et al.*, 2022; Wang *et al.*, 2023).

3.1.3 Ciclo de replicação

A adequada interação do vírus com os receptores das células do hospedeiro é a condição inicial para a infecção viral. O receptor para a infecção de células humanas e suínas por *Senecavirus A* é o receptor 1 da toxina do antraz (ANTXR1), também conhecido como marcador endotelial tumoral 8 (TEM8), e consiste em uma glicoproteína transmembrana na superfície celular. Este receptor está envolvido na fixação, internalização e desnudamento (Miles *et al.*, 2017; Chen *et al.*, 2022). Além do ANTXR1, estudos sugerem a participação de outras moléculas, como ácido siálico, na infecção por SVA (Liu *et al.*, 2013).

Após a ligação com o receptor ANTXR1, o virion é internalizado na célula hospedeira por uma via endocítica – macropinocitose não mediada por receptor ou vias de colesterol mediadas por receptor (Hou *et al.*, 2022; Jia *et al.*, 2022). A maioria dos picornavírus se liga aos seus receptores em uma depressão da superfície viral chamada de “cânion”, mas essa estrutura está ausente no SVA devido ao maior comprimento de suas alças expostas à superfície. A microscopia crioeletrônica do SVA indica que o ANTXR1 pode se ligar ao vírus pelos *loops* VP2-DF e VP1-CD, causando mudanças estruturais na alça VP1-GH e na alça VP3-GH, que atenuam as interações entre protômeros e desestabilizam a montagem do capsídeo (Cao *et al.*, 2018).

Após a internalização, ocorre o desnudamento do vírus e liberação do RNA viral no citoplasma da célula hospedeira. O mecanismo do desnudamento de SVA ainda não está esclarecido (Wang *et al.*, 2023). Para que o genoma de SVA seja liberado no citosol é necessário que a tradução seja iniciada pela maquinaria celular do hospedeiro. As moléculas de RNA produzidas podem ser usadas como molde na próxima tradução ou replicação, ou podem ser montadas em partículas virais maduras. A montagem e maturação dos vírions ainda não foram compreendidas (Kinobe *et al.*, 2022 *apud* Wang *et al.*, 2023). A transmissão ocorre célula a célula, mediante auxílio de exossomos, semelhante ao que ocorre com outros picornavirus (Zhang *et al.*, 2019b; Xu *et al.*, 2020).

3.1.4 Epidemiologia

Quadros de doença vesicular associados à infecção por *Senecavirus A* já foram reportados no Brasil (Vannucci *et al.*, 2015), Canadá (Pasma *et al.*, 2008), Chile (Bennett *et al.*, 2022), China (Wu *et al.*, 2017), Colômbia (Sun *et al.*, 2017), Estados Unidos (Singh *et al.*, 2012), México (Navarro-Lopez *et al.*, 2023), Tailândia (Saeng-chuto *et al.*, 2018) e Vietnã (Arzt *et al.*, 2019).

No Brasil, os primeiros casos da doença foram reportados em 2014 e 2015, nos estados de Minas Gerais (setembro e novembro/2014), Goiás (outubro/2014), Paraná (novembro/2014), São Paulo (dezembro/2014), Mato Grosso (fevereiro/2015), Mato Grosso do Sul (março/2015), Rio Grande do Sul (março/2015) e Santa Catarina (maio/2015) (Linhares *et al.*, 2015). No segundo semestre de 2018, novos surtos de doença vesicular foram notificados nos mesmos estados, com exceção do Mato Grosso do Sul, e a confirmação laboratorial de SVA fez com que estes casos fossem relacionados à segunda onda da doença no país (Vieira *et al.*, 2022). A terceira onda ocorreu em maio de 2020, com relatos de doença vesicular por SVA em Santa Catarina e no Paraná (Vieira *et al.*, 2022). A infecção por SVA é considerada endêmica em algumas regiões do país (Brasil, 2020).

Dados de soroprevalência são escassos no Brasil e no mundo. Nos Estados Unidos, estudo de Houston *et al.* (2019) realizado com amostras de propriedades sem histórico prévio de SVA encontrou soroprevalência de 42,7% em granjas de recria e terminação e 75,8% em granjas de matrizes reprodutoras. Já no estudo de Preis (2022), também em granjas norte-americanas, o resultado foi de 7,4% em granjas de recria e terminação e 17,3% em granjas de matrizes, discrepância que o autor buscou justificar a partir das particularidades das granjas analisadas nos dois estudos, período de coleta de amostras, interpretação de exames sorológicos, dentre outros aspectos. Mais da metade das províncias chinesas relataram casos de SVA, mas não existem dados concretos quanto à soroprevalência (Liu *et al.*, 2020). A prevalência da doença no Brasil também permanece desconhecida (Rieger *et al.*, 2023), sendo que o dado mais recente, referente a estudo conduzido com amostras de rebanhos de suínos do sul do Brasil, no ano de 2020, é de que 36,5% das 263 amostras avaliadas foram positivas para SVA, sendo 89 amostras de porcos sintomáticos e 7 de animais assintomáticos (Vieira *et al.*, 2022).

Quanto à susceptibilidade, até o momento, apenas os suínos foram identificados como hospedeiros de *Senecavirus A*, mas poucos trabalhos investigaram a possibilidade de infecção de outras espécies domésticas ou selvagens (Alexandersen *et al.*, 2019). A identificação de baixos títulos de anticorpos neutralizantes contra o vírus no soro de bovinos domésticos

(Knowles *et al.*, 2006), bem como o isolamento de SVA a partir de suabes orais de búfalos apresentando úlceras orais e claudicação (Zhou *et al.*, 2021), justificaram a realização de estudo para investigar a susceptibilidade de *Bos taurus* à infecção experimental por SVA (Buckley *et al.*, 2022). Os resultados apontam que, apesar de demonstrada susceptibilidade de células bovinas ao SVA nos testes *in vitro*, as análises *in vivo* indicam que não há replicação eficaz do vírus nos bovinos domésticos. Assim, até o momento, foi descartada a hipótese de importância epidemiológica da espécie na infecção por SVA. Ressalta-se que não existem indícios de que a infecção por SVA tenha potencial zoonótico (Knowles *et al.*, 2006; Alexandersen *et al.*, 2019).

O RNA de *Senecavirus A* também já foi detectado e isolado em camundongos e moscas. Em camundongos, o isolamento de SVA viável ocorreu a partir de amostras de fezes e de intestino delgado desses animais (Joshi *et al.*, 2016b). Quanto às moscas domésticas, estudo demonstrou a viabilidade prolongada do vírus e a possibilidade de este ser transportado pelas moscas por longas distâncias e por pelo menos 48 horas (Turner *et al.*, 2022). Estes estudos sugerem a participação de moscas e roedores na transmissão mecânica do vírus, o que torna o controle de pragas uma ferramenta essencial na dinâmica da infecção.

O vírus é transmitido, além da forma mecânica mencionada, principalmente pelo contato direto entre animais infectados e susceptíveis, a partir de secreções nasais e orais, fluido das vesículas e fezes (Joshi *et al.*, 2016b). A transmissão indireta do vírus a partir do ambiente e de fômites contaminados, semelhante ao que já foi descrito para outros patógenos, também já foi confirmada, inclusive com isolamento de vírus infeccioso de suabes de botas e macacões de pessoas em contato com animais infectados (Preis, 2022) e confirmação da infecção pelo consumo de alimentos contendo partículas virais (Caserta *et al.*, 2022). A transmissão vertical do vírus pelas matrizes foi associada à imunossupressão induzida por estresse (Maggioli *et al.*, 2019). A eliminação persistente e prolongada de SVA no sêmen também já foi demonstrada (Sturos *et al.*, 2022).

A detecção do RNA viral em amostras de animais infectados perdura por cerca de 28 dias, mas o período para isolamento de vírus viável varia com o tipo de amostra: 7 dias nas secreções nasais, 14 dias nas fezes e 21 dias nas secreções orais. Assim, o tempo de excreção viral é superior à duração dos sinais clínicos, que é em média de 14 dias (Maggioli *et al.*, 2018). Ressalta-se que estudos recentes sugerem a possibilidade de infecção persistente em alguns animais afetados, com detecção de SVA infeccioso nas tonsilas até 60 dias após a infecção (Maggioli *et al.*, 2019).

A taxa de morbidade é variável conforme a categoria afetada: até 70% em neonatos, 0,5-5% em leitões desmamados; 5-40% em suínos de terminação e reprodutores; e até 90% em

porcas. Em relação à mortalidade, esta é expressiva em neonatos (até 70%, relatada por Vannucci *et al.*, 2015), mas pouco significativa (0,2%) nas demais categorias (Leme *et al.*, 2017).

3.1.5 Patogenia

Após a infecção do animal susceptível, o *Senecavirus A* realiza sua replicação inicial nas tonsilas, conforme detalhado no item 3.1.3 – Ciclo de replicação. Entre um e dez dias ocorre viremia, e o SVA alcança seus sítios de predileção, como o focinho, lábios, banda coronária e espaços interdigitais. O período de incubação é curto, entre três e cinco dias, quando então o vírus induz a sintomatologia aguda e autolimitada caracterizada por áreas eritematosas que evoluem para vesículas de tamanho variável, que se rompem após cerca de 5 dias e progridem para erosão dérmica e ulceração, estando completamente cicatrizadas em aproximadamente duas semanas. Outros sinais clínicos frequentemente observados são claudicação, letargia e anorexia (Leme *et al.*, 2017; Houston *et al.*, 2020).

Apesar de a resolução clínica ser rápida, destaca-se novamente que a eliminação viral persiste por até 28 dias, com detecção de vírus infeccioso nas secreções nasais, fezes e secreções orais por até 7, 14 ou 21 dias, respectivamente; e que há indícios de animais com infecção persistente, o que deve ser levado em consideração devido ao risco de transmissão (Maggioli *et al.*, 2018 e 2019).

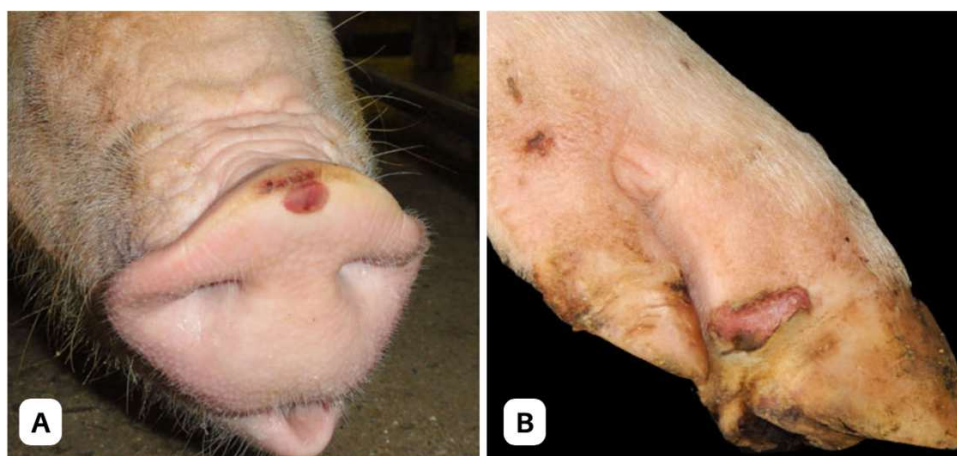


Figura 3 - Adaptada de Segalés *et al.*, 2017. Porcas naturalmente infectadas por SVA, com focinho apresentando pele ulcerada após ruptura de vesícula (A); e casco com ulceração da banda coronária (B).

Macroscopicamente, observam-se áreas de descamação, erosões, vesículas e ulcerações, principalmente nos lábios, focinho e membros distais. Ao nível histológico, já foram descritas dermatite da banda coronária, com separação da derme e epiderme e formação de fenda

contendo queratinócitos necróticos, células inflamatórias, hemorragia e edema. Além disso, foram observadas alterações em outros tecidos, como hiperplasia linfóide leve a moderada nas tonsilas, baço e gânglios linfáticos; e áreas de atelectasia multifocal leve e congestão difusa, com acúmulo de células de defesa, nos pulmões (Joshi *et al.*, 2016a; Leme *et al.*, 2017).

Nos neonatos, a infecção por SVA resulta em quadro clínico multissistêmico, sendo comumente detectada sintomatologia de doença vesicular nas porcas. Os sinais clínicos observados em leitões neonatos afetados são fraqueza, letargia, salivação excessiva, diarreia e sinais neurológicos, com até 70% dos animais evoluindo à óbito. Este quadro ficou conhecido como Perdas Neonatais Epidêmicas Transientes (PNET), devido à alta mortalidade e à resolução do quadro em até 12 dias (Vannucci *et al.*, 2015; Leme *et al.*, 2016).

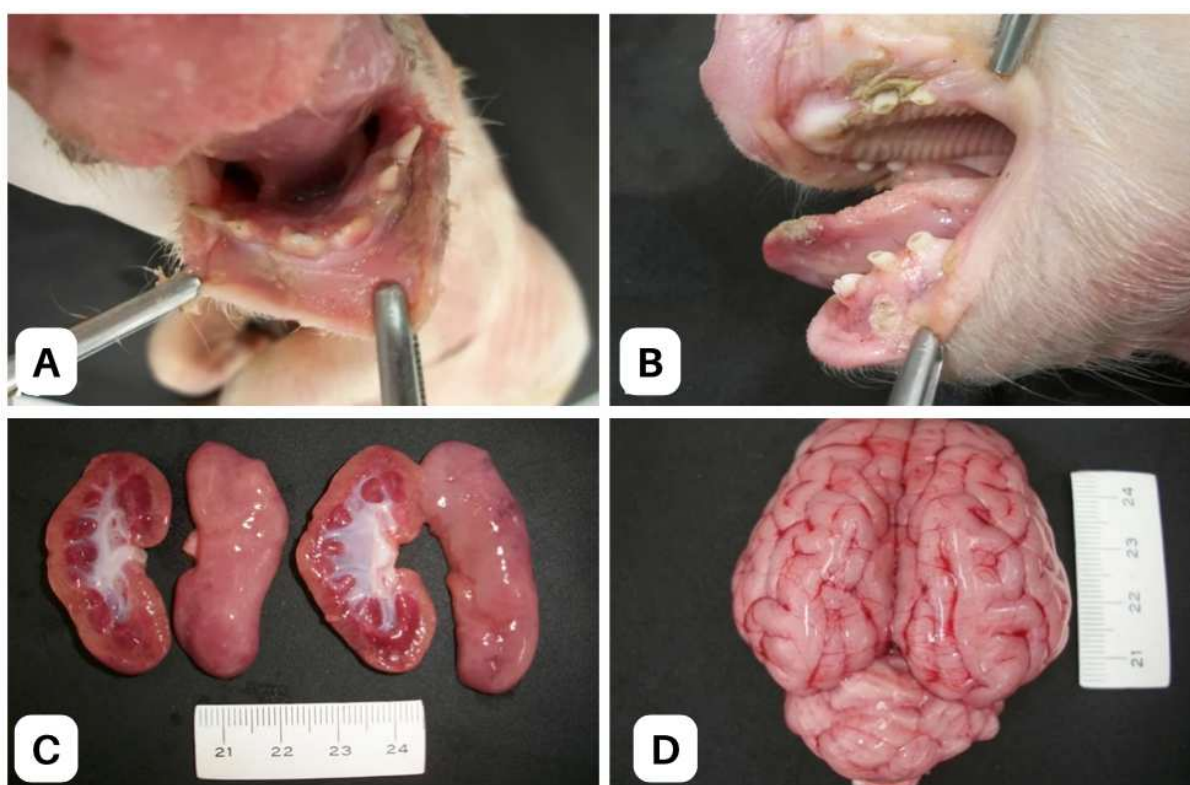


Figura 4 - Adaptada de Leme *et al.*, 2016. Lesões identificadas em leitões infectados com *Senecavirus A*: gengivite multifocal (A), glossite e gengivite (B), nefrite hemorrágica (C) e congestão dos vasos meníngeos (D).

As lesões macroscópicas já relatadas em casos de PNET foram impressões fracas de costelas na superfície pleural dos pulmões, edema pulmonar, congestão de vasos meníngeos, hemorragias petequiais nos rins, glossite, gengivite, lesões ulcerativas na banda coronária e fezes líquidas, indicativas de diarreia, no intestino delgado. Como alterações histológicas, foram reportadas pneumonia intersticial, depleção linfóide no baço, tonsilas ou gânglios

linfáticos; degeneração hidrópica (balonosa) do epitélio da bexiga urinária e dos ureteres; encefalite linfoplasmocítica e atrofia das vilosidades intestinais nos leitões que apresentavam diarreia (Leme *et al.*, 2016).

3.1.6 Imunidade

Diversas moléculas estão envolvidas no reconhecimento do RNA de *Senecavirus A* e no desenvolvimento da resposta imune à infecção. Dentre elas, merecem destaque os receptores do tipo Toll 3 e 7 (*toll-like receptors* – TLRs 3 e 7), o gene I induzível por ácido retinóico (RIG-I) e a proteína 5 associada à diferenciação de melanoma (MDA—5) (Zhao *et al.*, 2022).

Após o reconhecimento viral são produzidos títulos importantes de anticorpos neutralizantes, que podem ser detectados a partir do quinto dia pós infecção, havendo expressiva correlação destes com o isotipo IgM (figura 5). Estudo de Maggioli e colaboradores (2018) verificou que a produção de IgM e IgG foi direcionada contra as três proteínas estruturais localizadas na face externa do capsídeo (VP1, VP2 e VP3), entretanto as respostas mais robustas foram associadas à VP2 e VP3. Neste trabalho, as imunoglobulinas do tipo IgM não foram detectadas após 21 dias pós-infecção, assim como a resposta de IgG contra VP1 e VP3. Quanto à VP2, anticorpos IgG foram detectados até 35 dias pós infecção. Já no trabalho de Gimenez-Lirola e colaboradores (2016), a detecção de IgG anti-VP1 se estendeu por seis semanas após surto de SVA.

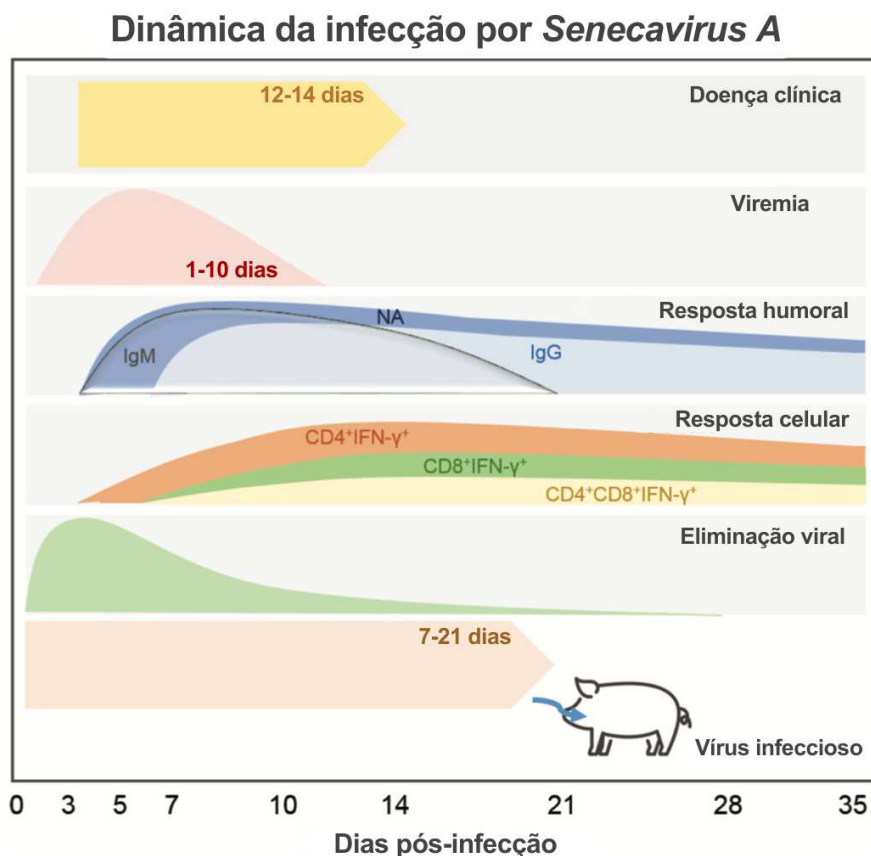


Figura 5 - Adaptada de Maggioli *et al.*, 2018. Os sinais clínicos causados pela infecção por SVA, como presença de vesículas, letargia e claudicação, surgem de três a cinco dias após a infecção, e perduram por até 14 dias. A viremia é curta, de 1 a 10 dias. Anticorpos neutralizantes (NA) são detectados do quarto ou quinto dia até pelo menos 35 dias após a infecção. A resposta celular pode ser detectada a partir de 7 dias pós-infecção e é intensificada até o 14º dia. A eliminação viral ocorre por até 28 dias, com isolamento de vírus infeccioso de secreções nasais, fezes e secreções orais até os dias 7, 14 e 21 após a infecção, respectivamente.

A presença do vírus estimula também a imunidade celular, com forte resposta de células T CD4⁺ a partir do sétimo dia pós infecção. A partir do décimo dia, são também detectadas células T CD8⁺ e células T duplamente positivas (T CD4⁺ CD8⁺). A identificação de células T em amostras coletadas no 35º dia após a infecção, em teste de reestimulação *in vitro*, indica uma provável indução de células T de memória (Maggioli *et al.*, 2018; Houston *et al.*, 2020).

As células T CD4⁺ se diferenciam em células Th1, essenciais na defesa contra patógenos intracelulares, sob estímulo de IFN- γ secretado por células *natural killer* (NK). Estas células Th1 irão também secretar IFN- γ , que por sua vez irá intensificar a diferenciação neste tipo celular e inibir o desenvolvimento das células Th2 e Th17; estimular a expressão de proteínas que contribuem para a apresentação de antígenos e ativação de células T; e ativar macrófagos e outras células para que ocorra a morte de patógenos fagocitados e a produção de citocinas amplificadoras da resposta imune (Abbas *et al.*, 2022).

Conforme brevemente mencionado, o RIG-I e o MDA-5 estão associados ao reconhecimento viral, e após se ligarem ao RNA do vírus desencadeiam uma cascata de sinalização que ativa fatores de transcrição, como o fator regulador do IFN γ (IRF3) e o fator nuclear κ B (NF- κ B). Um dos resultados da cascata é a produção de IFN do tipo I, importante para a expansão e maturação da resposta imune adaptativa (Zhao *et al.*, 2022).

Embora estudos mais aprofundados sobre os aspectos imunológicos da infecção por *Senecavirus A* sejam escassos, alguns mecanismos de evasão do sistema imune utilizados pelo vírus vêm sendo elucidados. Dentre eles, a inibição da produção de IFN do tipo I a partir da regulação da expressão do IRF3 e do IRF7 pela proteína 3C; a degradação de RIG-1 associada às proteínas 2C e 3C; a atuação das proteínas 2B e 3C no comprometimento da funcionalidade da proteína de sinalização antiviral mitocondrial (MAVS), que medeia a ativação de NF- κ B e IRFs e induz a produção de IFN; a inibição da autofagia, a partir da clivagem do receptor sequestossomo-1 (SQSTM1) pela proteína 3C e da interação da proteína 2AB com a ubiquitina ligase MARCHF8; e a suposta inibição, pela proteína 3C, da formação de grânulos de estresse reguladores da tradução do mRNA, que requer maiores investigações (Houston *et al.*, 2020; Zhao *et al.*, 2022).

3.1.7 Diagnóstico

A infecção por *Senecavirus A* é clinicamente indistinguível de outras doenças vesiculares clássicas como febre aftosa, estomatite vesicular, doença vesicular dos suínos e o exantema vesicular dos suínos (as duas últimas exóticas no Brasil) (Brasil, 2020). Desta forma, aliar o histórico da granja, os sinais clínicos e o diagnóstico laboratorial é imprescindível para associar um quadro de doença vesicular a um agente específico. Devido à similaridade clínica com febre aftosa, deve ser feita a notificação da suspeita de doença vesicular ao serviço veterinário oficial, para que a investigação clínica e epidemiológica seja iniciada e, caso identificados sinais clínicos compatíveis, seja feita a interdição da propriedade e a coleta e remessa de amostras para diagnóstico no LFDA-MG (Becker *et al.*, 2020).

Métodos diretos e indiretos estão disponíveis para detecção laboratorial de SVA. Dentre os métodos diretos, o isolamento viral a partir de suabes e esfregaços de vesículas foi bem sucedido em diferentes linhagens de células, evidenciado pelo efeito citopático característico de arredondamento e lise celular. Como exemplos de células usadas para o isolamento de SVA tem-se células de retinoblastos fetais (PER.C6) (Venkataraman *et al.*, 2008), células de rim de suíno (PK-15) (Leme *et al.*, 2015), células de testículo de suíno (ST) (Zhang *et al.*, 2015) e células de carcinoma pulmonar humano (NCI-H1299) (Yang *et al.*, 2012). Devido à

complexidade, gasto de materiais e tempo envolvidos na execução da técnica de cultivo celular, sua utilização é mais comum no desenvolvimento de pesquisas do que na rotina diagnóstica.

A RT-PCR (reação de transcriptase reversa seguida de reação em cadeia da polimerase) é outro método de diagnóstico direto, que se destaca por sua alta sensibilidade e especificidade e por ser uma técnica que já está amplamente disponível na maioria dos laboratórios. Diferentes alvos virais já foram utilizados neste teste diagnóstico, como o gene da polimerase 3D, em RT-PCR convencional (Laguardia-Nascimento *et al.*, 2016) e em tempo real (RT-qPCR) (Laguardia-Nascimento *et al.*, 2016; Fowler *et al.*, 2017); e regiões conservadas da VP1, em RT-qPCR e também em ensaio de nested RT-PCR (Dall Agnol *et al.*, 2017; Feronato *et al.*, 2018). A região 3D também foi utilizada por Pinheiro-de-Oliveira *et al.* (2019) na técnica de RT-PCR digital em gotas (RT-ddPCR). Quanto às amostras utilizadas nos testes de RT-PCR, o RNA viral já foi detectado nas fezes, fluido oral, soro e tecidos diversos, como fígado, intestino delgado, cérebro, pulmão, linfonodos, tonsilas, dentre outros (Gimenez-Lirola *et al.*, 2016; Resende *et al.*, 2017; Preis, 2022). Ressalta-se que já foi detectado RNA de SVA em amostras de animais sem sinais clínicos, demonstrando a ocorrência de quadros subclínicos da infecção (Bracht *et al.*, 2016).

A técnica de amplificação isotérmica mediada por *loop* de transcrição reversa (RT-LAMP) também já foi desenvolvida para diagnóstico de SVA, utilizando amostras de vesículas e tecidos (Zeng *et al.*, 2018). A combinação da técnica de RT-LAMP com dispositivo de fluxo lateral, descrita por Li e colaboradores (2019), é uma opção promissora para diagnóstico a campo, assim como o RT-PCR isotérmico isolado (RT-iiPCR) associado a dispositivo portátil POCKIT™ desenvolvido por Zhang *et al.* (2019a). Os dois métodos portáteis foram desenvolvidos utilizando soro como amostra, mas no trabalho de Zhang e colaboradores (2019a) além do soro foram testados suabes vesiculares, suabes de tonsila, fluidos orais e suabes retais. Considerando o curto período de viremia, de até 10 dias, entende-se que a detecção a campo deve utilizar prioritariamente espécimes clínicos alternativos ao soro, como os testados por Zhang *et al.* (2019a), os quais permitem a detecção viral em um intervalo maior de tempo de infecção, conforme destacado no tópico 3.1.4 (Maggioli *et al.*, 2018 e 2019).


A detecção de antígeno ou ácido nucleico de SVA nos tecidos pode ser realizada por imunohistoquímica (IHQ) e hibridização *in situ* (ISH). Estudo de Leme e colaboradores (2016) utilizou anticorpos monoclonais para detecção de antígeno de SVA em tecidos fetais, observando marcação imunohistoquímica na língua, gengiva, pelve renal, bexiga, cérebro, cerebelo e intestino delgado. Sondas específicas de RNA, projetadas contra regiões do gene VP1 e aplicadas na técnica de ISH por Resende e colaboradores (2017), demonstraram a

presença de SVA em lesão vesicular, coração, linfonodos, baço, pulmão, fígado, cólon, intestino delgado, língua e tonsila.

Outro método que vêm sendo utilizado para *Senecavirus A* é o sequenciamento, inclusive de amostras brasileiras (Vannucci *et al.*, 2015; Vieira *et al.*, 2022). Trata-se de uma tecnologia apta a identificar com precisão o agente etiológico e fornecer informações valiosas a nível molecular, mas devido ao custo mais elevado de equipamentos e reagentes e à necessidade de etapas de análise de bioinformática, o sequenciamento é uma técnica mais restrita à área de pesquisa. A metodologia de sequenciamento será detalhada no item 3.2.1.

A importância da detecção direta do vírus é incontestável, mas, conforme destacado por Gimenez-Lirola e colaboradores (2016), variações quanto ao tempo de excreção do vírus e à carga viral nas diferentes amostras biológicas podem comprometer o diagnóstico. Assim, métodos indiretos assumem relevância, especialmente para monitorar a circulação viral silenciosa em granjas endêmicas, nas quais não é comum a manifestação de sinais clínicos, mas o risco de transmissão da doença perdura, especialmente no contexto de introdução de animais susceptíveis à infecção (Houston *et al.*, 2020).

Dentre os métodos indiretos, o *indirect enzyme-linked immunosorbent assay* (iELISA) é comumente utilizado. Gimenez-Lirola e colaboradores (2016) desenvolveram iELISA a partir de proteína VP1 recombinante, obtendo sensibilidade de 93% e especificidade de 99%. Estudo de Dvorak *et al.* (2017) descreve o uso das proteínas estruturais VP1, VP2 e VP3 em iELISA, com melhores resultados associados à VP2 (sensibilidade de 94,2% e especificidade de 89,7%). Já no trabalho de Rieger *et al.* (2023), iELISA com VP2 recombinante apresentou sensibilidade de 74,4% e especificidade de 100%. Epítipo linear identificado na proteína VP2 foi utilizado como antígeno em iELISA desenvolvido por Ma *et al.* (2022), no qual a sensibilidade e a especificidade foram de 91,13% e 91,17%, respectivamente; com concordância de 90,41% com o kit comercial *Swinecheck*[®] *SVA bELISA* (BioVet Inc., Canadá). Outro grupo de pesquisa desenvolveu iELISA com a proteína 3AB recombinante, obtendo sensibilidade de 91,30% e especificidade de 92,59% (Yan *et al.*, 2023).

Testes de ELISA competitivos (cELISA) também já foram descritos. O cELISA desenvolvido por Yang *et al.* (2012) alcançou 100% de especificidade ao utilizar antígeno de SVA inativado e anticorpos monoclonais. Já no estudo de Goolia *et al.* (2017), foram observadas especificidade de 98,2% e sensibilidade de 96,9%. Bai e colaboradores (2021) desenvolveram cELISA a partir de partículas semelhantes a vírus (*virus-like particles* – VLPs), encontrando especificidade de 100%, sensibilidade de 94% e concordância de 90% com o kit comercial *As*  *Kit* (BioStone Animal Health, EUA).

Ma e colaboradores (2023) descrevem o desenvolvimento de cELISA para detecção de anticorpos neutralizantes contra SVA, no qual foi observada sensibilidade de 98,11% e especificidade de 100%.

Outros testes indiretos que já foram descritos na literatura para diagnóstico de SVA são a imunofluorescência indireta (Dvorak *et al.*, 2017; Goolia *et al.*, 2017), a soroneutralização (Goolia *et al.*, 2017) e a imunoperoxidase em monocamada de células (IPMC) (Rocha, 2022).

Ressalta-se que, no Brasil, o laboratório oficial (LFDA-MG) preconiza as técnicas de soroneutralização, com células NCI-H1299, e de RT-qPCR, com o gene da polimerase 3D como alvo, para diagnóstico de SVA (Brasil, 2022).

3.1.8 Prevenção e controle

As medidas preventivas contra *Senecavirus A* coincidem com as medidas comuns para biossegurança nas granjas, como controle da entrada de veículos, pessoas, alimentos, animais e equipamentos – componentes que podem estar associados ao transporte do vírus para o interior da granja; controle de moscas e roedores – dada a provável atuação como vetores mecânicos do vírus; aquisição de animais de reposição em granjas livres do agente; respeito ao tempo de quarentena; limpeza e desinfecção adequada das instalações; realização de vazio sanitário; monitoramento soropidemiológico do rebanho; dentre outras medidas (Houston *et al.*, 2020).

Especificamente quanto ao uso de desinfetantes, Leme e colaboradores (2017) sugerem a adoção do protocolo de desinfecção recomendado para o vírus da febre aftosa, que inclui hidróxido de sódio (2%), carbonato de sódio (4%), ácido cítrico (0,2%), ácido acético (2%), hipoclorito de sódio (3%), peroximonossulfato de potássio/cloreto de sódio (1%) e dióxido de cloro. No que diz respeito ao SVA, estudos foram conduzidos para avaliar a eficácia de diferentes desinfetantes. Singh e colaboradores (2017) observaram que o hipoclorito de sódio (5,25%) na diluição 1:20 inativou de forma eficaz ($\geq 99,99\%$) o vírus a 25°C, em até 15 minutos, em superfícies como alumínio, borracha e plástico. Em aço inoxidável e cimento, a inativação viral foi de 99,97% e 99,98%, respectivamente. A 4°C a eficácia foi similar à descrita para 25°C, em todas as superfícies, com exceção da borracha, na qual a inativação foi de 99,91% após 15 minutos. O desempenho de outros desinfetantes testados, como compostos a base de fenol e de amônia quaternária, foi inferior. Estudo de Hole *et al.* (2017) avaliou a eficácia de produtos comerciais à base de peróxido de hidrogênio (1:20) e monopersulfato de potássio (1%) contra SVA, sendo demonstrada a inativação viral após dez minutos de exposição aos produtos.

Além dos desinfetantes mencionados, a luz ultravioleta C (UVC) foi avaliada como alternativa para inativação de SVA em superfícies comumente encontradas em granjas, como papelão, tecido e plástico. A exposição à UVC foi considerada eficaz apenas em materiais plásticos livres de matéria orgânica, não sendo uma opção satisfatória para inativação do vírus em plásticos contaminados com elementos orgânicos ou em superfícies como papelão e tecido (Ruston *et al.*, 2021).

No Brasil, no ano de 2021, foi licenciada no MAPA a Senevac (registro nº 10.443/2021), uma vacina comercial inativada contra *Senecavirus A*, produzida pela empresa Inata Produtos Biológicos (Inata, [s.d.]). A recomendação da bula é a aplicação intramuscular de duas doses de 2mL, com intervalo de 21 dias, sendo a primeira aos 40 dias de idade. Apesar de a vacina estar licenciada há dois anos, até o momento não foram disponibilizados estudos a respeito da sua eficácia.

A nível experimental, outras opções vacinais vêm sendo desenvolvidas. Yang e colaboradores (2018) desenvolveram vacina inativada contra SVA, utilizando Montanide™ ISA 206 (Seppic, França) como adjuvante em suínos. Os animais vacinados desenvolveram altos títulos de anticorpos neutralizantes e não apresentaram sinais clínicos após o desafio. Estudo de Mu e colaboradores (2020) comparou a eficácia de vacina inativada com formulação baseada em partículas pseudovirais (VLPs), composta pelas proteínas do capsídeo VP0, VP1 e VP3, demonstrando que as duas opções vacinais induziram proteção semelhante em suínos desafiados com o vírus. Li *et al.* (2021) utilizaram uma amostra de SVA isolada de granja chinesa em 2018 como antígeno em vacina inativada, associada ao adjuvante Montanide™ ISA 201. Leitões de 100 dias imunizados com essa formulação por via intranasal produziram altos títulos de anticorpos neutralizantes, não desenvolveram lesões e não excretaram o vírus nas fezes após desafio com vírus homólogo. Outra vacina inativada, elaborada com amostra viral contendo mutações de aminoácidos em VP1, VP3 e 3C demonstrou segurança e imunogenicidade promissora em camundongos, evitando lesões e presença viral nos tecidos (Li *et al.*, 2022). Amostra viral chinesa de 2017 inativada com etilenoimina binária (BEI) e β -propiolactona (BPL) apresentou bons resultados em suínos de terminação, com ambos os adjuvantes testados (Montanide™ ISA 201 e Montanide™ IMG 1313) (Liu *et al.*, 2022b). Buckley e Lager (2022) avaliaram a eficácia de vacina inativada em leitões desmamados e porcas, bem como a transferência de imunidade passiva, com resultados promissores.

Outros tipos de vacina, além da inativada, já foram testados experimentalmente contra SVA. A vacina viva atenuada desenvolvida a partir de tecnologia recombinante e genética reversa por Sharma *et al.* (2019) resultou em proteção de leitões contra o desafio heterólogo de

SVA, constatada pela ausência de sinais clínicos e pela redução da viremia, excreção viral e carga viral nos tecidos. Tao e colaboradores (2023) desenvolveram vacina recombinante utilizando vírus da pseudorraiva como vetor para expressão da proteína VP2 de SVA. Testes conduzidos em camundongos induziram altos níveis de anticorpos neutralizantes contra os dois vírus. A predição *in silico* de epítomos de células B foi utilizada por Zhang *et al.* (2023) para identificar alvos vacinais no SVA. Os epítomos detectados nas proteínas VP1 e VP2 foram utilizados como antígeno em vacina recombinante multi-epítomo, que induziu anticorpos neutralizantes e proporcionou 80% de proteção em leitões imunizados e desafiados com vírus homólogo.

A utilização de ferramentas de bioinformática, semelhante ao descrito no trabalho de Zhang *et al.* (2023), vem trazendo perspectivas promissoras para o desenvolvimento de vacinas e métodos diagnósticos. Tendo em vista o potencial desta metodologia e sua relação direta com o presente trabalho, conceitos importantes serão aprofundados a seguir.

3.2 Ferramentas de bioinformática

3.2.1 Métodos de sequenciamento genômico

O sequenciamento genômico consiste na identificação, na ordem correta, da sequência de nucleotídeos de moléculas de DNA ou RNA. Assim, o sequenciamento vem se consolidando como uma ferramenta importante para identificação de patógenos e para estudos de expressão gênica, diversidade genética, relações evolutivas, dentre outros (Fietto e Maciel, 2015).

O primeiro método de sequenciamento foi descrito em 1977 por Sanger e colaboradores. Nesta tecnologia, que ficou conhecida como sequenciamento Sanger, a sequência complementar à fita molde é obtida a partir da adição de dideoxynucleotídeos de terminação de cadeia marcados – anteriormente com compostos radioativos, e depois fluorescentes –, e da separação e análise dos fragmentos por eletroforese em gel para determinação da sequência (Sanger *et al.*, 1977; Hu *et al.*, 2021).

O rendimento limitado da metodologia Sanger motivou o desenvolvimento de novas tecnologias, culminando no sequenciamento de nova geração (NGS). Como exemplos, tem-se as plataformas 454, *Illumina* e *Ion Torrent*, tecnologias de sequenciamento de leitura curta, de até 600nts/leitura (*short-read sequencing*), que seguem etapas de preparação de biblioteca, sequenciamento e análise de dados (Hu *et al.*, 2021).

A preparação da biblioteca de DNA envolve inicialmente a obtenção de moldes correspondentes às moléculas de interesse – por meio de PCR ou captura de hibridização, por

exemplo. Outras etapas são a fragmentação, que utiliza métodos físicos, enzimáticos ou químicos para cortar o DNA na faixa ideal de tamanho conforme cada plataforma; o reparo final, que prepara as bibliotecas para a ligação aos adaptadores; a ligação propriamente dita, com sequências que serão reconhecidas pela plataforma de sequenciamento como ponto de início e fim da leitura; e a seleção do tamanho a partir de esferas ou eletroforética, para enriquecer os fragmentos de DNA dentro de uma faixa de tamanho definida e remover contaminantes. Após estas etapas, é realizado o sequenciamento conforme o princípio de cada plataforma. Quando se trata de patógenos que possuem RNA como material genético, a preparação da amostra inclui o isolamento do RNA, o enriquecimento do RNA alvo e a transcrição reversa do RNA em DNA complementar (cDNA). Ao final, o produto da amplificação é processado para sequenciamento (Hu *et al.*, 2021).

A etapa de sequenciamento nas plataformas NGS ocorre a partir da incorporação de nucleotídeos na fita de DNA (sequenciamento por síntese), que são detectados por diferentes tecnologias conforme a plataforma utilizada. No sequenciador 454 (*Roche*), NGS de primeira geração, a construção da sequência decorre da liberação de pirofosfato enquanto são adicionados os nucleotídeos, resultando na emissão de luz, que é detectada e interpretada pelo equipamento. Já nas plataformas NGS *Illumina* (figura 6) e *Ion Torrent*, de segunda geração, a incorporação de nucleotídeos é identificada pela fluorescência emitida; ou pela liberação de íon hidrogênio; respectivamente (Rothberg & Leamon, 2008; Hu *et al.*, 2021).

O uso de plataformas para sequenciamento de trechos curtos de DNA apresenta desvantagens, como a dificuldade de identificar variantes estruturais, grandes inserções ou deleções (*indels*), elementos repetitivos e diferentes mutações no mesmo genoma. Estes inconvenientes assumem importância especialmente quando se trata de genomas virais, nos quais diferentes fontes de variação coexistem, como mutações pontuais, recombinação e *indels* (Hu *et al.*, 2021; Quer *et al.*, 2022).

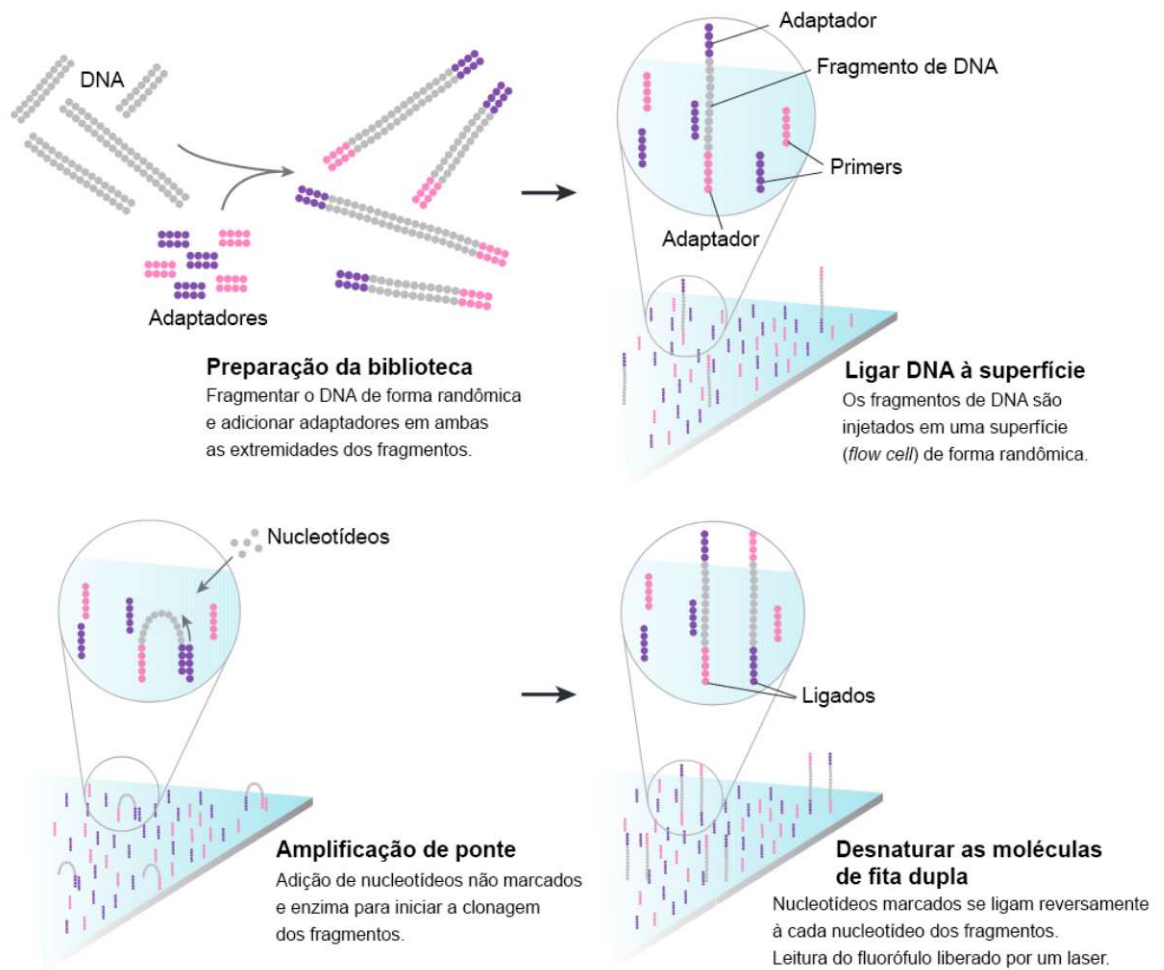


Figura 6 - Adaptada de Mardis (2008), por Balbinot (2020). Representação esquemática da tecnologia *Illumina*. A preparação da biblioteca, passo inicial do protocolo, envolve a fragmentação do DNA de forma randômica e adição de adaptadores nas extremidades 3' e 5'. Em seguida, os fragmentos são injetados em uma lâmina (*flow cell*) e clonados por amplificação de ponte. Posteriormente, nucleotídeos marcados são sequencialmente adicionados para que se liguem reversamente aos nucleotídeos dos fragmentos, com liberação de fluoróforo e leitura deste pelo laser do equipamento.

Os desafios referentes ao sequenciamento de genomas virais levaram ao desenvolvimento de soluções alternativas, como o sequenciamento a partir de primers aleatórios, do enriquecimento de regiões alvo, do uso de amplicons ou de plataformas de leitura longa.

A primeira alternativa, conhecida como sequenciamento metagenômico, envolve o uso de primers aleatórios, que consistem em oligonucleotídeos curtos de DNA de fita simples (6-9 nucleotídeos) em todas as combinações possíveis de bases. Desta forma, os iniciadores podem se ligar a qualquer sequência alvo de DNA ou RNA, que poderá ser amplificada em PCR e

posteriormente sequenciada em plataforma NGS após a composição de biblioteca. Como vantagens, este método exclui a necessidade de conhecimento prévio do genoma para o desenho de primers ou sondas e permite a identificação dos patógenos em amostras variadas (plasma, soro, fezes, esgoto, etc.). Como desvantagens, genomas minoritários podem ser perdidos, por exemplo quando há baixa carga viral, uma vez que o genoma do hospedeiro e de microrganismos comensais também será amplificado. Com isso, o custo computacional para recuperação de dados de interesse é alto. Uma solução seria concentrar o agente infeccioso antes do sequenciamento, a partir do uso de anticorpos, filtração, ultracentrifugação ou depleção de material do hospedeiro com métodos de filtração de DNase (Quer *et al.*, 2022).

Outro método utilizado para genomas virais é o sequenciamento a partir do enriquecimento de regiões alvo. Neste, pequenas sondas sobrepostas de RNA ou DNA são projetadas de forma complementar às sequências de referência do patógeno. As sondas são ligadas à uma fase sólida, como esferas (*beads*) magnéticas, e usadas para capturar as sequências complementares. Após a ligação a adaptadores, as sequências são enriquecidas com poucos ciclos de PCR e então sequenciadas. As principais vantagens deste método são a obtenção de sequências de microrganismos específicos dentre outros genomas que podem estar presentes na mesma amostra, a partir das regiões alvo selecionadas pelas sondas; e o menor custo para sequenciar o genoma inteiro, pois a contaminação por presença de material genético de outros organismos é reduzida. Como desvantagens, o desenho das sondas requer conhecimento sobre o genoma do microrganismo, inviabilizando o uso da técnica para descoberta de novos patógenos; é exigido pré-tratamento específico das amostras; e a cobertura do sequenciamento é em regra baixa (Quer *et al.*, 2022).

O sequenciamento mediante uso de amplicons também foi descrito como estratégia para sequenciamento de genomas virais, e consiste em projetar primers específicos para amplificar parcial ou completamente um genoma por PCR. O produto da PCR é então utilizado para o sequenciamento. A sensibilidade alta, a amplificação específica de agentes infecciosos com reduzida contaminação por outros genomas, o bom custo-benefício, e a boa cobertura mesmo diante de baixa carga viral são as principais vantagens da técnica. Entretanto, os resultados são dependentes da qualidade e especificidade dos primers, e quanto mais ciclos necessários na PCR maior o número de artefatos que podem comprometer a exatidão dos dados (Quer *et al.*, 2022).

Por fim, o sequenciamento de leitura longa corresponde ao sequenciamento de terceira geração (TGS), representado por plataformas como *Nanopore* (Oxford, Reino Unido) e *Quantapore* (Roche, EUA). De forma simplificada, a tecnologia consiste na passagem de uma

molécula de DNA ou RNA de fita simples por nanoporos e na análise das interrupções da corrente iônica, que cada nucleotídeo promove de forma distinta, para determinação da sequência. Como essa metodologia supera os inconvenientes relatados para o sequenciamento de leitura curta, como a dificuldade de identificar *indels*, sua utilização está associada a melhores resultados para o sequenciamento de genomas virais. O maior percentual de erros sistemáticos devido a falhas na consistência dos sinais elétricos, as dificuldades para obtenção de leituras sobrepostas, a grande quantidade de material genético necessária para o sequenciamento direto e o baixo rendimento são as principais desvantagens citadas (Hu *et al.*, 2021; Quer *et al.*, 2022)

A tabela 1 destaca as principais características das três gerações de sequenciamento:

Tabela 1 - Vantagens e desvantagens das principais plataformas de sequenciamento

(Adaptada de Quer *et al.*, 2022 e Pervez *et al.*, 2022)

	Primeira geração	Segunda geração	Terceira geração
Exemplos de plataformas	Sanger	<i>Illumina, Ion Torrent</i>	<i>Nanopore, PacBio</i>
Vantagens	Alta acurácia e auxílio na validação de resultados de NGS	Alto rendimento com baixa taxa de erros; custo menor.	Leituras longas permitem reconstrução mais precisa do genoma; custo ainda menor; sem necessidade de amplificação por PCR.
Desvantagens	Alto custo e baixo rendimento	Leituras curtas pouco adequadas para montagem e caracterização de regiões genômicas complexas/repetitivas; necessidade de amplificação por PCR.	Alta taxa de erros (até 20%) dificulta a identificação de substituições, inserções e deleções reais.

3.2.1.1 Sequenciamento de amostras de *Senecavirus A*

Diferentes grupos de pesquisa sequenciaram, parcial ou totalmente, amostras de *Senecavirus A*. Na plataforma NCBI *Virus* (www.ncbi.nlm.nih.gov/labs/virus/vssi/#/; acesso em 08/01/2024), 421 sequências de nucleotídeos estão disponíveis, enquanto a plataforma BV-BRC (www.bv-brc.org/view/Virus/10239; acesso em 08/01/2024), que abriga o ViPR, reúne 431 sequências. Em ambas as plataformas há apenas uma sequência de referência, denominada SVV-001 (acesso *Genbank* n. NC_011349). Ao aplicar filtros à busca realizada na plataforma BV-BRC é possível verificar que a maioria das sequências foi obtida pela metodologia de Sanger. Outras metodologias que constam nos resultados da pesquisa são 454, *Illumina*, *IonTorrent* e *Nanopore MinIon*.

3.2.2 Processamento dos dados do sequenciamento

Os dados obtidos no sequenciamento devem passar por etapas de controle de qualidade, filtragens, montagem, identificação taxonômica, validação e análise. Diferentes ferramentas estão disponíveis, e vários fluxogramas de trabalho já foram descritos (Cantalupo e Pipas, 2019).

Após a obtenção das leituras brutas, no formato *fastq*, estas devem passar por uma verificação de qualidade (Andrews, 2014; Cantalupo e Pipas, 2019). A ferramenta FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) é uma das opções para esta etapa, e traz as seguintes informações (Babraham Bioinformatics, [s.d.], a):

- Estatísticas básicas (figura 7): traz informações como o nome do arquivo analisado, o tipo de arquivo, o padrão de codificação, o total de sequências, número de sequências marcadas como de má qualidade e filtradas da análise, o comprimento das sequências e o conteúdo GC (% GC) médio de todas as bases. Não existem advertências ou falhas associadas às estatísticas básicas.

Basic Statistics

Measure	Value
Filename	good_sequence_short.txt
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	250000
Total Bases	10 Mbp
Sequences flagged as poor quality	0
Sequence length	40
%GC	45

Figura 7 - Babraham Bioinformatics, [s.d.],b. Estatísticas básicas apresentadas em resultado hipotético de sequenciamento *Illumina*.

- Qualidade da sequência por base (figura 8): o gráfico mostra uma visão geral da faixa de valores de qualidade de todas as bases, em cada posição. A linha vermelha central é o valor mediano, a caixa amarela representa o intervalo interquartil (25-75%), os limites superior e inferior do *boxplot* representam os pontos de 10% e 90% e a linha azul representa a qualidade média. O eixo y mostra os índices de qualidade, e o fundo

do gráfico mostra áreas coloridas de acordo com a qualidade: a área verde indica qualidade muito boa, a amarela indica qualidade razoável e a vermelha está associada a uma qualidade ruim. É esperado que a qualidade decaia ao longo da execução do sequenciamento, por limitações inerentes à própria técnica, como o número de pares de bases coberto pelo kit ou o consumo de reagentes. Este parâmetro emite advertência (símbolo !) se o quartil inferior de qualquer base for inferior a 10, ou se a mediana de qualquer base for inferior a 25. Será emitido um erro (símbolo X) caso o quartil inferior de qualquer base for inferior a 5 ou se a mediana de qualquer base for inferior a 20. O motivo mais comum para advertências e falhas é a degradação esperada da qualidade ao longo de longas execuções, o que pode ser resolvido com um corte de qualidade (truncamento das leituras com base na qualidade média) e remoção do conteúdo de adaptadores. Uma outra possível fonte de erros é uma perda de qualidade no início da execução, com posterior recuperação, o que pode estar relacionado com problemas transitórios como a presença de bolhas em uma célula de fluxo. Neste caso não se recomenda o corte, pois pode causar a perda de regiões importantes da sequência. Coberturas baixas em determinado intervalo base também podem ocasionar advertências neste módulo.

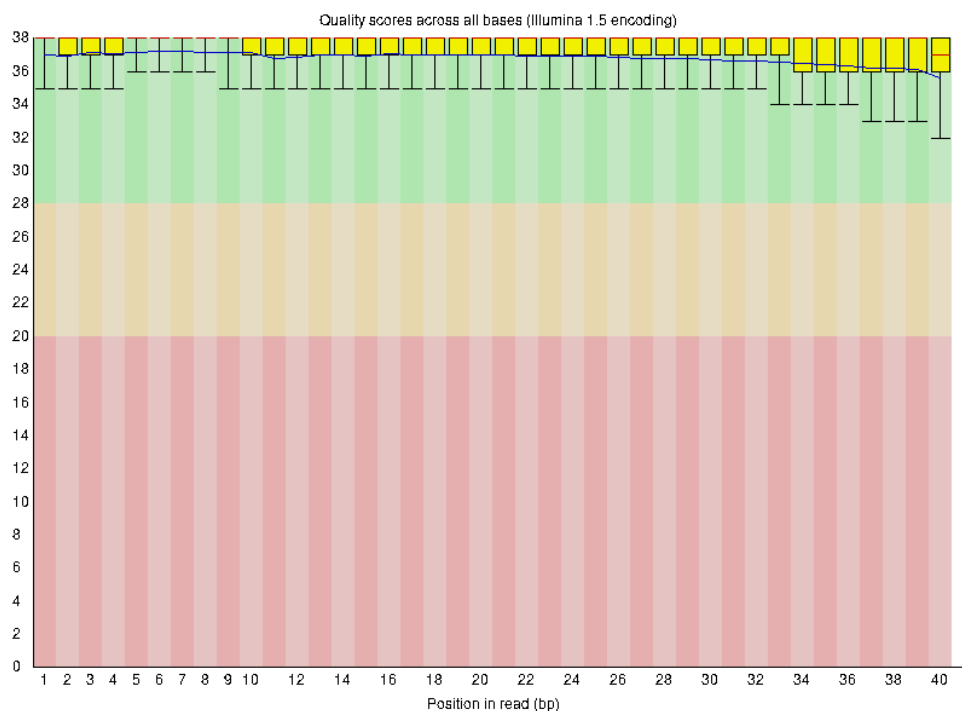


Figura 8 - Babraham Bioinformatics, [s.d.], b. Qualidade de sequência por base, em bom resultado de sequenciamento *Illumina*, obtido a partir de dados hipotéticos. A área verde (pontuação superior a 28) indica qualidade muito boa, a amarela (pontuação entre 20 e 28) indica qualidade razoável e a vermelha (pontuação inferior a 20) está associada a uma qualidade ruim.

- Qualidade da sequência por bloco (figura 9): este gráfico está restrito a casos de utilização de biblioteca *Illumina* que retenham os identificadores de sequência originais. Os identificadores codificam o bloco da célula de fluxo associado a cada leitura. O gráfico permite observar os índices de qualidade de cada bloco em todas as suas bases, a fim de verificar se houve perda de qualidade em apenas uma parte da célula de fluxo, representando com cores frias as posições em que a qualidade estava igual ou acima da média daquela base na execução, e com cores quentes os blocos com qualidades piores do que outros para aquela base. Quanto mais o gráfico apresentar tons azuis, melhor o resultado. A queda de qualidade por bloco pode estar associada a problemas transitórios, como bolhas, ou problemas permanentes, como detritos na célula de fluxo. Uma advertência é emitida se algum bloco mostrar uma pontuação média de *Phred* (valores no eixo y) mais que duas vezes inferior à média dos outros blocos. Já um erro será reportado caso a diferença de pontuação seja maior que cinco vezes. Erros que afetem pequeno número de blocos em regra são ignorados, enquanto erros persistentes por vários ciclos demandam investigações.

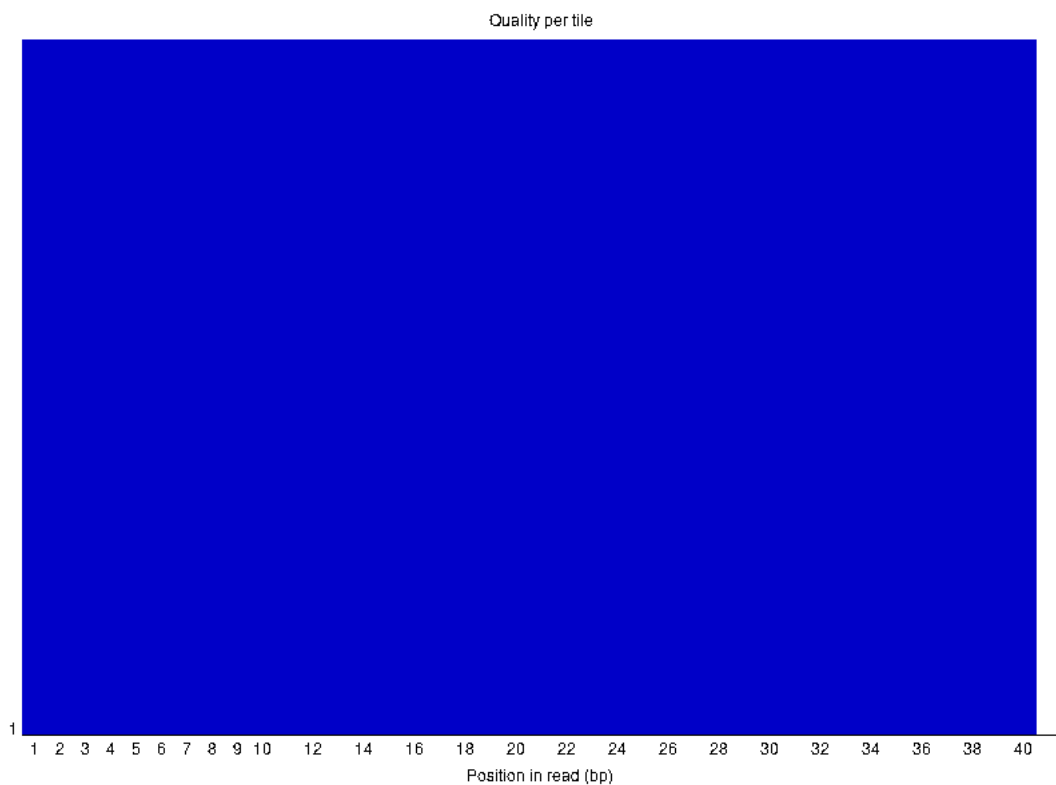


Figura 9 - Babraham Bioinformatics, [s.d.], b. Qualidade de sequência por bloco, em bom resultado de sequenciamento *Illumina* (hipotético), evidenciado pela cor azul presente no gráfico como um todo.

- *Scores* de qualidade por sequência (figura 10): este gráfico permite avaliar se um subconjunto de sequências possui valores de qualidade universalmente baixos. É aceitável que um ou outro subconjunto possua qualidade inferior, por falhas pontuais na execução, mas estas devem representar uma pequena porcentagem do total de sequências. Uma advertência será emitida se a qualidade média observada com maior frequência for inferior a 27, o que equivale a uma taxa de erro de 0,2%. Já um aviso de erro será gerado se a qualidade média mais frequente for inferior a 20, ou seja, se a taxa de erro alcançar 1%. A critério do pesquisador, podem ser feitos cortes de qualidade, especialmente tratando-se de corridas longas.

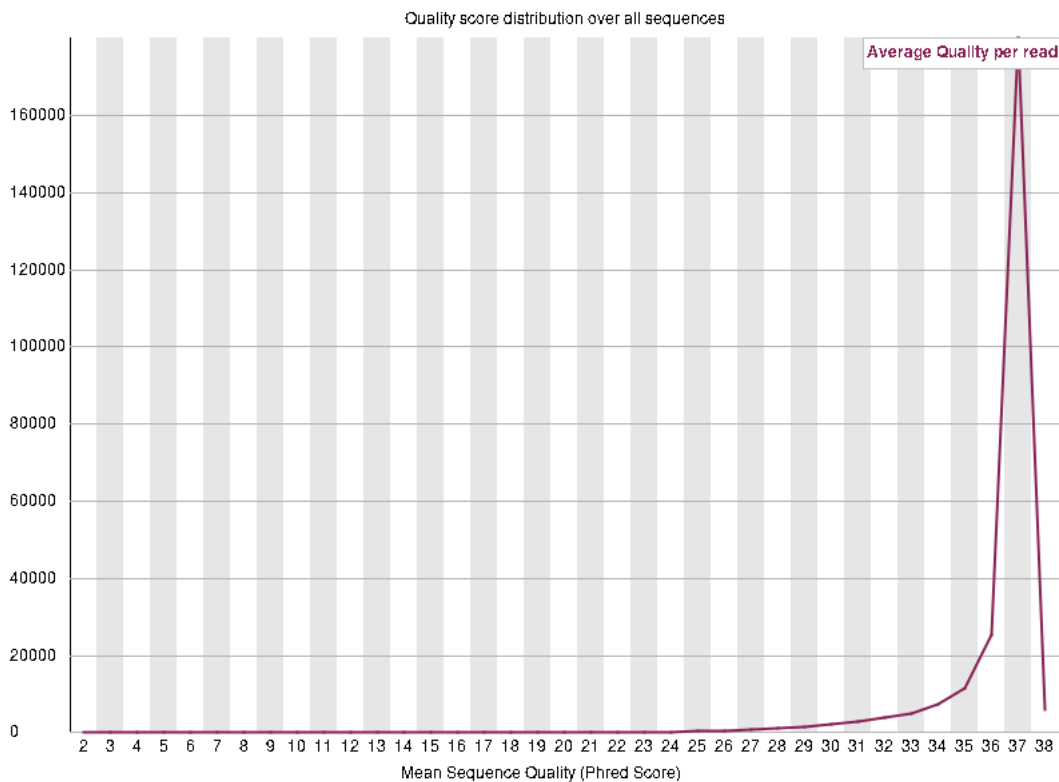


Figura 10 - Babraham Bioinformatics, [s.d.], b. *Score* de qualidade por sequência, em resultado hipotético de sequenciamento *Illumina*, onde se vê uma boa média da qualidade por leitura.

- Conteúdo da sequência por base (figura 11): espera-se que a quantidade relativa de cada base reflita a quantidade total das bases no genoma e, portanto, elas não devem apresentar desequilíbrios marcantes. Assim, quanto mais paralelas as linhas neste gráfico, melhor o resultado. Algumas bibliotecas podem produzir uma composição de bases tendenciosa no início da leitura, como as de RNA-Seq e aquelas fragmentadas por transposase – nestes casos, há um viés intrínseco nas posições em

que as leituras começam, que não parecem comprometer a análise posterior, mas resultam em alertas de advertência ou erro. A advertência será emitida caso a diferença entre A e T, ou G e C, seja maior que 10% em qualquer posição; enquanto o erro será emitido se a diferença entre A e T, ou G e C, for maior que 20% em qualquer posição. As principais causas de falhas associadas a este módulo, além da já mencionada fragmentação tendenciosa por transposase ou por bibliotecas RNA-Seq, são a presença de sequências sobrerepresentadas, identificadas por dímeros adaptadores ou rRNA; bibliotecas que possuem algum viés, como aquelas tratadas com bissulfito de sódio, que convertem a maior parte das citosinas em timinas; e bibliotecas que foram cortadas por adaptadores de forma agressiva, causando desvios repentinos na composição de bases observada.

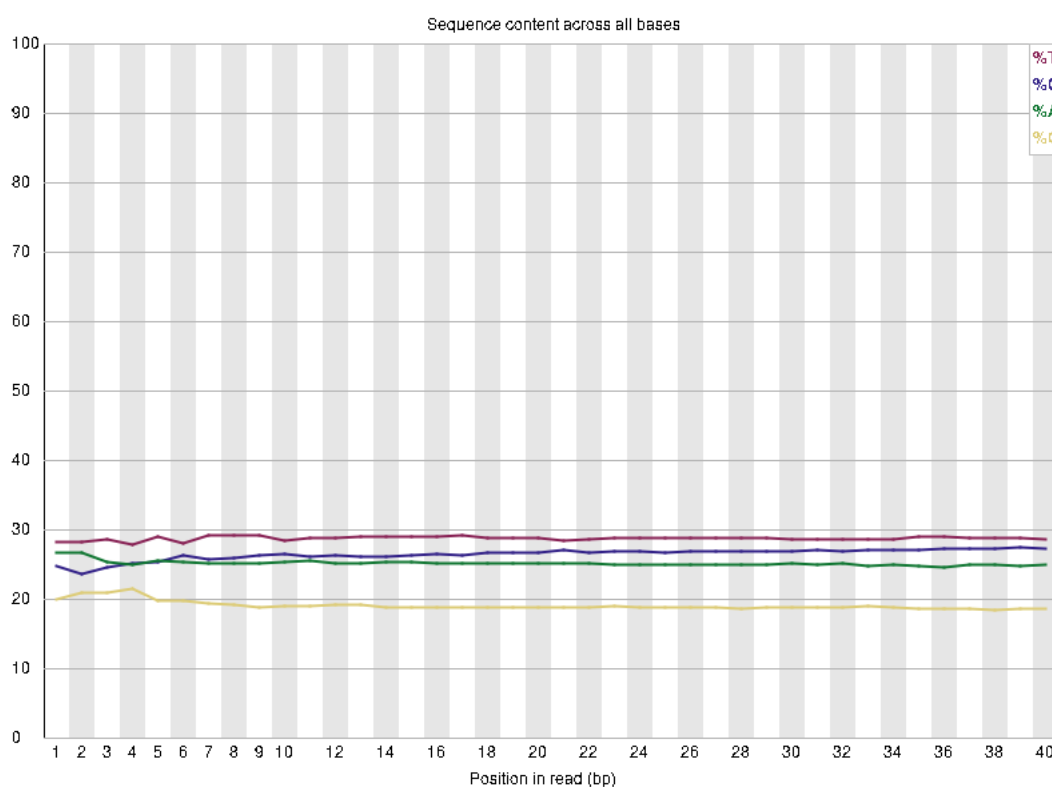


Figura 11 - Babraham Bioinformatics, [s.d.], b. Conteúdo de sequência por base, em resultado hipotético de sequenciamento *Illumina*. As linhas paralelas indicam um equilíbrio das bases no genoma, o que é considerado um bom resultado.

- Conteúdo GC por sequência (figura 12): o gráfico apresenta uma curva teórica de distribuição normal do conteúdo GC (em azul), permitindo a comparação com o conteúdo GC da sequência. Uma distribuição anormal pode indicar contaminação da biblioteca ou que há subconjuntos tendenciosos. Picos agudos em uma distribuição

suave geralmente estão associados a um contaminante específico, como um adaptador, que pode facilmente ser identificado; enquanto picos amplos podem representar contaminação por uma espécie diferente. Uma advertência será gerada se a soma dos desvios da distribuição normal representar mais de 15% das leituras. Caso a soma atinja mais de 30% das leituras, será emitido um sinal de erro.

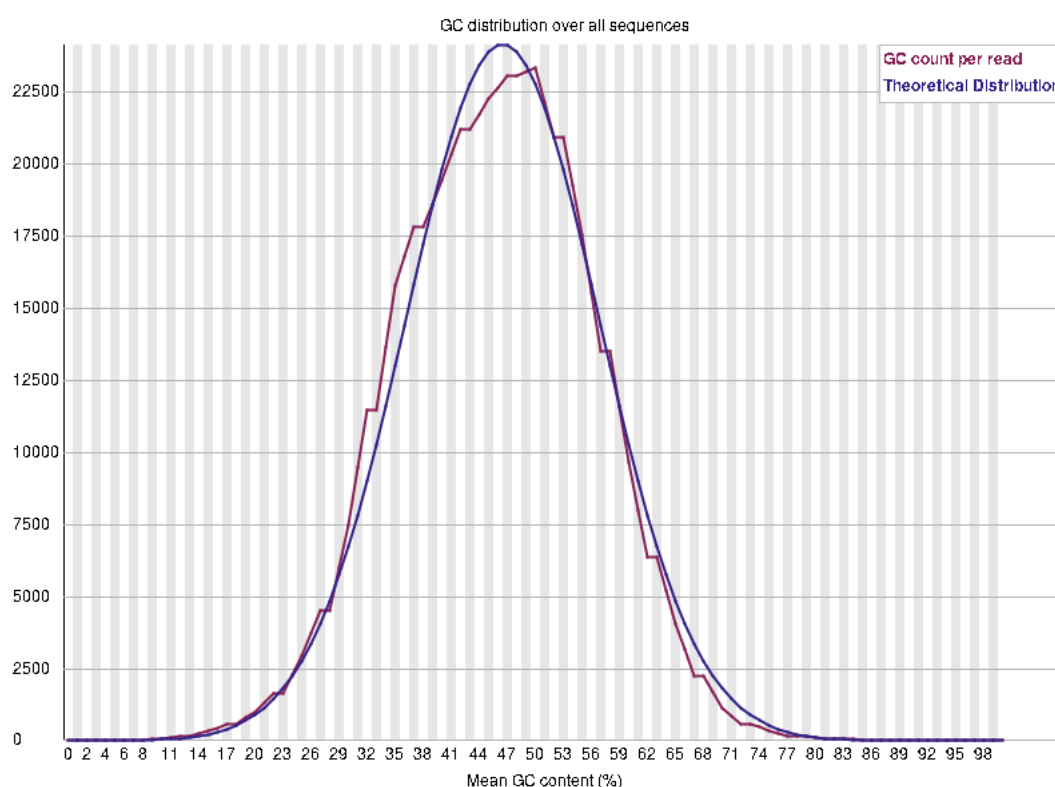


Figura 12 - Babraham Bioinformatics, [s.d.], b. Conteúdo GC por base, em resultado hipotético de sequenciamento *Illumina*. O gráfico mostra que as leituras seguem distribuição normal, indicando um bom resultado.

- Conteúdo N por base (figura 13): quando a plataforma de sequenciamento não consegue determinar qual o nucleotídeo presente em determinada posição com confiabilidade suficiente, é adicionada uma base N ao invés de uma base convencional. Será emitida uma advertência ou um erro se alguma posição apresentar um conteúdo N superior a 5% ou 20%, respectivamente. A razão mais comum para o excesso de Ns é uma perda geral de qualidade ou desvios devido a bibliotecas tendenciosas, que afetam a confiabilidade das leituras.

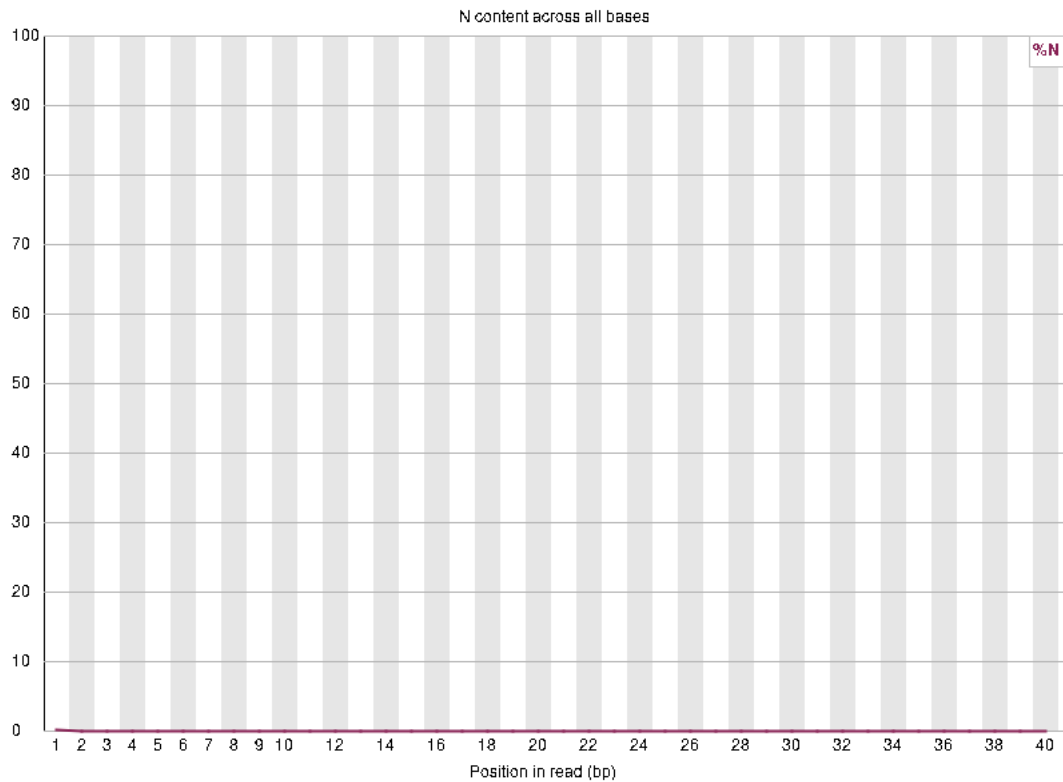


Figura 13 - Babraham Bioinformatics, [s.d.], b. Conteúdo N por base, em resultado hipotético de sequenciamento *Illumina*. A ausência de conteúdo N indica um resultado positivo, compatível com uma boa cobertura.

- Distribuição de comprimento das sequências (figura 14): algumas plataformas de sequenciamento geram fragmentos de comprimento uniforme, enquanto outras fornecem leituras de tamanhos variados. No primeiro caso, será gerado um gráfico simples, mostrando um pico apenas em determinado tamanho; ao passo que no segundo caso o gráfico irá apontar as quantidades relativas de cada tamanho de fragmento. Uma advertência será gerada caso os fragmentos não tenham o mesmo tamanho, o que como dito poder ser apenas uma característica da plataforma de sequenciamento utilizada; e um erro será reportado se alguma sequência possuir comprimento zero.

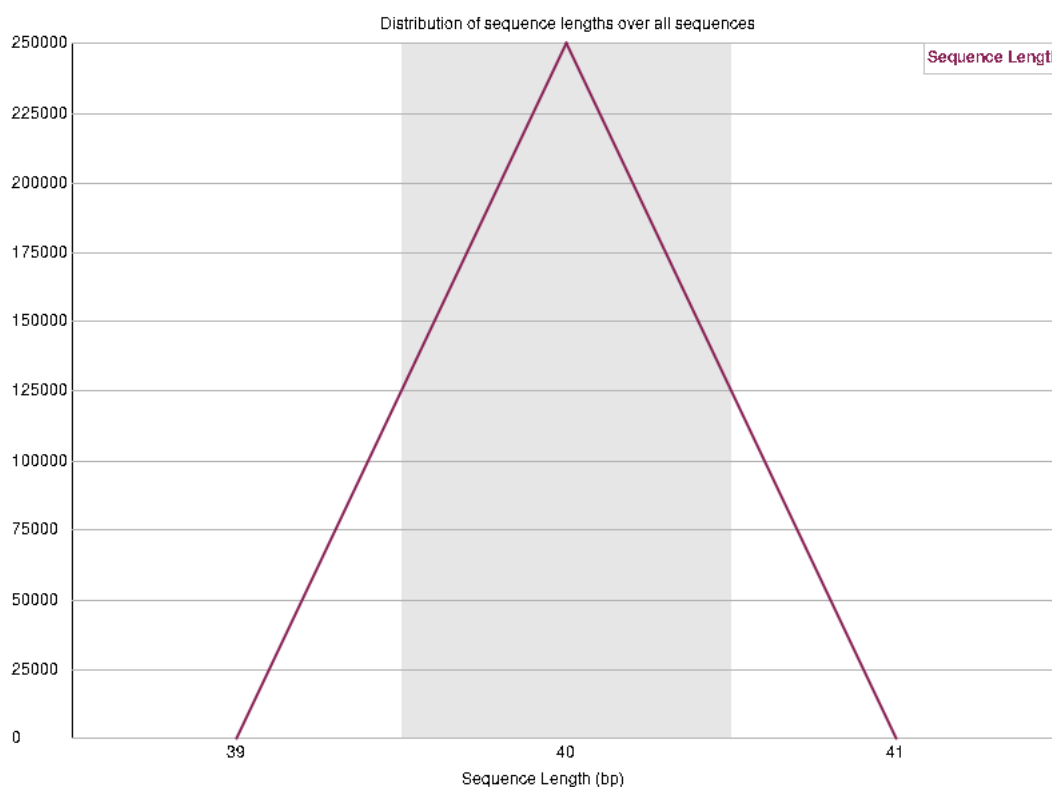


Figura 14 - Babraham Bioinformatics, [s.d.], b. Distribuição do comprimento das sequências, em resultado hipotético de sequenciamento *Illumina*. O resultado mostrado é característico desta plataforma, a qual fornece fragmentos de tamanho uniforme.

- Sequências duplicadas (figura 15): em regra, a maioria das sequências ocorre apenas uma vez no conjunto final da biblioteca. Neste módulo, o grau de duplicação é calculado e plotado em um gráfico que mostra o número relativo de sequências com diferentes graus de duplicação: a linha azul indica o conjunto completo de sequências e como seus níveis de duplicação são distribuídos; e a linha vermelha mostra a distribuição dos níveis de duplicação das sequências deduplicadas (ou seja, sequências que foram processadas para eliminar dados repetidos e reduzir o conjunto de dados analisados). O esperado para este gráfico, para as duas linhas, é que a maioria das sequências esteja reunida na extremidade esquerda do gráfico, ou seja, com nível baixo de duplicação, indicando boa cobertura. Um alto nível de duplicação, por outro lado, sugere uma tendência de enriquecimento, por artefatos de PCR, por exemplo. Níveis gerais de enriquecimento estão associados ao sequenciamento excessivo da biblioteca, levando ao achatamento de linhas. Quando o enriquecimento é mais específico a determinado subconjunto ou oriundo da presença de contaminantes, serão observados picos à direita, especialmente na linha azul, pois

constituem grande proporção da biblioteca original (quando consideradas as sequências deduplicadas, em vermelho, a proporção é insignificante). Quando os picos da linha azul persistem, grande parte das sequências está altamente duplicada, revelando problemas como por exemplo a presença de contaminantes diversos. Se sequências duplicadas representarem mais de 20% do total será emitida uma advertência; enquanto um erro será indicado quando o percentual de duplicações for acima de 50%. Uma advertência ou erro neste módulo é uma declaração de que foi esgotada a diversidade em pelo menos parte da biblioteca, com novo sequenciamento das mesmas sequências. Em alguns tipos de biblioteca, como nas de RNA-Seq, essa situação é comum, visto que para poder observar transcrições de baixa expressão, é comum sequenciar demasiadamente as transcrições de alta expressão, o que potencialmente resulta em um grande conjunto de duplicatas. Análise das regiões de maior ocorrência de duplicatas podem auxiliar na distinção entre sequências excessivas biológicas e duplicação técnica.

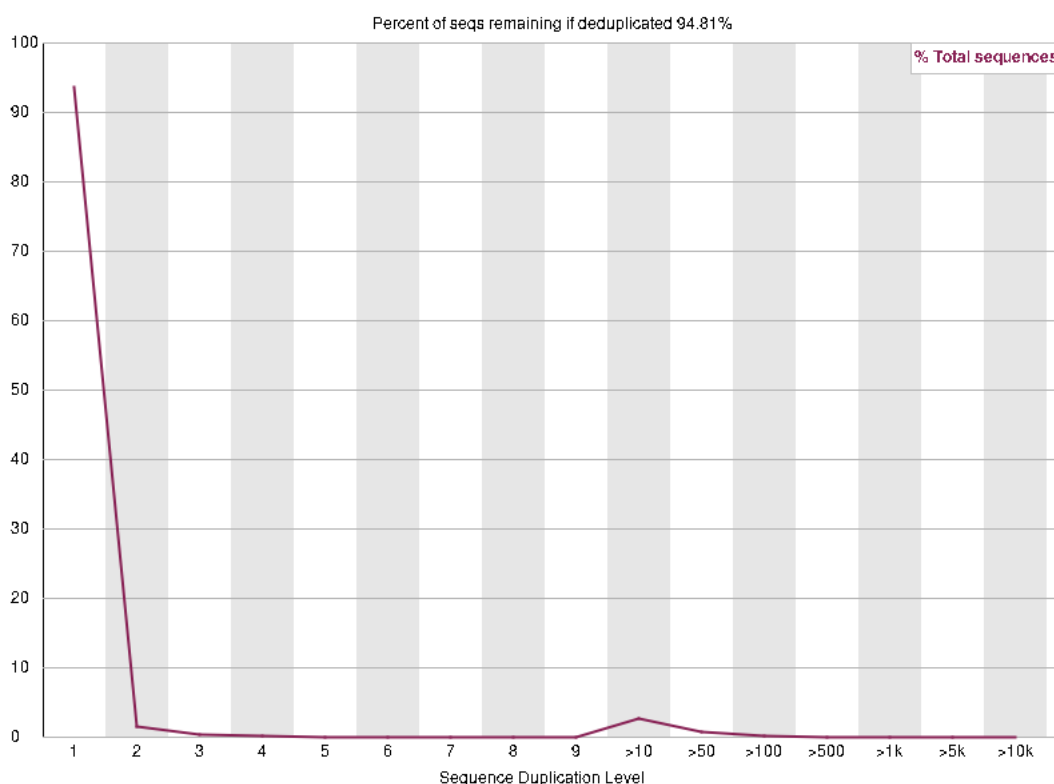


Figura 15 - Babraham Bioinformatics, [s.d.], b. Nível de sequências duplicadas, em resultado hipotético de sequenciamento *Illumina*. O resultado mostrado atende ao esperado, com a maioria das sequências reunidas no lado esquerdo do gráfico, indicando um nível de duplicação aceitável.

- Sequências super-representadas (figura 16): a presença de sequências super-representadas no conjunto pode indicar uma maior importância biológica destas, ou a presença de contaminantes, ou uma biblioteca de menor diversidade. Considera-se como super-representadas todas as sequências que representam mais de 0,1% do total. Para cada sequência super-representada, o programa procura correspondências em base de dados de contaminantes comuns, em uma tentativa de identificar a fonte da contaminação. Estes dados são apresentados na forma de tabela. Sequências que representem mais de 0,1% do total sinalizam uma advertência; e aquelas que representem mais de 1% do total geram um erro.



Figura 16 - Babraham Bioinformatics, [s.d.], b. Sequências super-representadas. Em um bom resultado de sequenciamento pela plataforma *Illumina*, é esperada a ausência de sequências super-representadas. No caso de estas existirem, são apresentadas na forma de tabela, com coluna indicando possível fonte de contaminação.

- Conteúdo do adaptador (figura 17): este módulo avalia a presença de subsequências de leitura de mesmo comprimento, buscando identificar adaptadores comumente utilizados nas plataformas de sequenciamento (adaptador universal *Illumina*, adaptador de RNA *Illumina*, transposase *Nextera* e adaptador de RNA SOLID, por exemplo). O gráfico mostra o percentual cumulativo da sequência de adaptadores em cada posição, de modo que a curva do gráfico vai aumentando ao longo da leitura. A advertência é emitida caso alguma sequência esteja presente em mais de 5% de todas as leituras; enquanto mais de 10% de presença resulta em erro. A remoção dos adaptadores deve ser feita antes da montagem do genoma para evitar interferências.

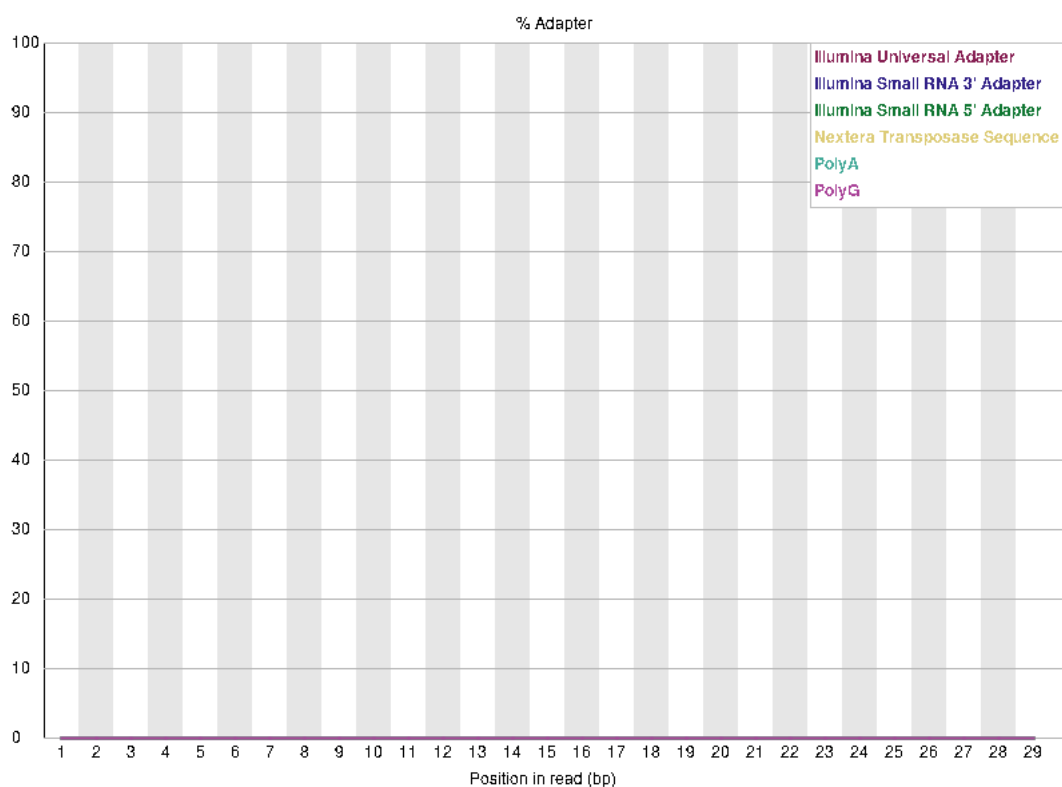


Figura 17 - Babraham Bioinformatics, [s.d.], b. Conteúdo de adaptadores. O resultado mostrado é o ideal, no qual não são identificadas sequências de adaptadores interferindo nos dados sequenciados.

- Conteúdo K-mer: K-mers são subsequências de mesmo comprimento (K) presentes em uma leitura. Os adaptadores, abordados no módulo anterior, são exemplos de K-mers. A análise parte do pressuposto de que qualquer fragmento de sequência não deve ter um viés posicional em uma biblioteca diversificada. O teste mede o número de cada 7-mer em cada posição, e utiliza um teste binomial para buscar desvios de cobertura posicionais. Quaisquer K-mers com enriquecimento posicional são relatados em uma tabela, e os seis mais frequentes aparecem plotados no gráfico. Caso haja um desequilíbrio de K-mers com valor de p binomial $< 0,01$, será emitida uma advertência. Valores de p binomial $< 10^{-5}$ geram um erro. Comumente sequências super-representadas aparecem enriquecidas no módulo K-mer, mesmo que não atinjam o patamar necessário para constarem no módulo correspondente a super-representação, aparecendo como picos acentuados em um único ponto da sequência. Quando há um bom resultado neste módulo, ou seja, quando não há identificação de K-mers com viés posicional, nenhum gráfico ou tabela é apresentado.

Programas como Prinseq (prinseq.sourceforge.net) podem ser utilizados para remover leituras de baixa qualidade e adaptadores, dentre outras funções (Schmieder e Edwards, 2011). Outra filtragem que pode ser aplicada é a remoção de conteúdo celular ou do hospedeiro (Cantalupo e Pipas, 2019).

A montagem do genoma é outra etapa necessária após a obtenção dos dados do sequenciamento, e consiste na ordenação dos *contigs*, ou seja, dos agrupamentos das leituras (*reads*) fornecidas pela plataforma de sequenciamento. Existem três principais formas de montagem: montagem *de novo*, montagem por referência ou mapeamento, e a combinação entre elas. A montagem *de novo* é baseada na tentativa de determinar conexões entre as leituras, ou seja, no alinhamento destas, e resulta em uma sequência consenso criada a partir de algoritmos. Uma vantagem desta metodologia é dispensar um genoma de referência, o que pode ser favorável para microrganismos recém estudados. O mapeamento por referência, por sua vez, ocorre a partir do alinhamento dos *contigs* com um genoma de referência, levando à obtenção de uma sequência similar a este. É uma alternativa que permite rapidamente identificar se os genes e proteínas esperados estão presentes, mas podem existir perdas de regiões importantes, que são descartadas por não constarem na referência. Quando não são identificadas determinadas regiões na referência, os montadores inserem espaços (*gaps*) no genoma em processo de montagem, representados por sequências Ns com tamanho superior a 10. A combinação dos métodos mencionados aumenta a precisão da montagem (Cantalupo e Pipas, 2019; Kremer, 2020).

A qualidade da montagem pode ser inferida pela proximidade do tamanho obtido do genoma com o esperado para a espécie e por índices como a cobertura, que reflete a quantidade de *reads* associadas a determinado fragmento de DNA. Uma cobertura de 10X, por exemplo, indica que cada nucleotídeo foi encontrado em pelo menos 10 *reads* naquele determinado genoma. A montagem de genomas obtidos por plataformas NGS de segunda geração, caracterizadas por leituras curtas, pode ser desafiadora do ponto de vista da cobertura, uma vez que comumente há dificuldade em sequenciar regiões repetitivas, que podem levar a lacunas no genoma devido à incerteza na montagem dessas regiões; ou regiões com conteúdo GC excessivamente alto ou baixo, que podem não ser tão bem sequenciadas comparado com regiões com conteúdo GC balanceado. Outro parâmetro utilizado é o N50, que revela o quanto do genoma é coberto por *contigs* grandes. Por exemplo, um valor de N50 igual a 100.000 indica que 50% das *reads* estão associadas a *contigs* de 100.000 bases ou maiores. Assim, quanto maior o valor de N50, melhor a qualidade da montagem (Staats *et al.*, 2014; Wang *et al.*, 2021).

Após a montagem do genoma é possível realizar o alinhamento, no qual sequências são organizadas em linhas e algoritmos específicos buscam a melhor correspondência entre elas. Quando necessário, são criados *gaps* entre os caracteres, de modo que ao final todas as sequências tenham o mesmo comprimento e seja mais fácil visualizar a similaridade entre elas. A identificação taxonômica pode ser obtida por meio de algoritmos baseados em alinhamento, como BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi) ou Bowtie2 (bio.sourceforge.net/bowtie2/) (Altschul *et al.*, 1990; Langmead e Salzberg, 2012; Cantalupo e Pipas, 2019). Após o alinhamento, a filogenia molecular pode ser utilizada para esclarecer relações evolutivas, processos biológicos, dinâmicas populacionais, distribuição geográfica, dentre outras possibilidades (Braun *et al.*, 2014; Staats *et al.*, 2014).

Outra etapa posterior à montagem é a anotação do genoma. Esta consiste na predição da função de genes e proteínas, permitindo que seja dada interpretação biológica aos dados obtidos. A anotação pode ser feita baseada em homologia ou por meio de abordagens *ab initio*, que são baseadas em modelos estatísticos (Ejigu e Jung, 2020). Um exemplo de programa recomendado para anotação de genomas virais é o BV-BRC (*Bacterial and Viral Bioinformatics Resource Center* - www.bv-brc.org), que engloba o ViPR (*Virus Pathogen Resource*) (Olson *et al.*, 2023). Para anotação de proteínas, o Uniprot (www.uniprot.org/) é comumente utilizado (The Uniprot Consortium *et al.*, 2023). A curadoria manual do genoma, processo pelo qual são utilizados programas para verificar e corrigir eventuais erros oriundos das etapas anteriores, pode aumentar a confiabilidade da montagem (McDonnell *et al.*, 2018; Howe *et al.*, 2021).

3.2.3 Análises de filogenia

A filogenética é a ciência que busca inferir sobre a história evolutiva dos organismos a partir das sequências de nucleotídeos ou aminoácidos. Seus resultados podem ser expressos graficamente a partir das chamadas árvores filogenéticas, que são formadas por estruturas conhecidas como nós, ramos, raiz e unidades taxonômicas operacionais (OTUs - *operational taxonomical units*), conforme representado na figura 18 (Braun *et al.*, 2014). A construção das árvores filogenéticas envolve o alinhamento das sequências, a escolha do modelo evolutivo e do método de construção e a validação das informações.

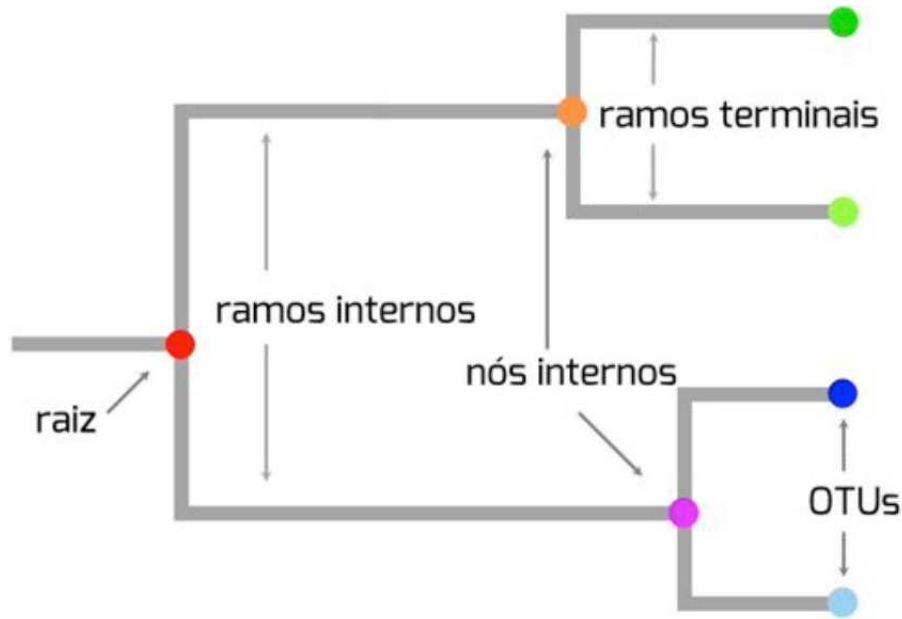


Figura 18 - Braun *et al.*, 2014. Representação esquemática das estruturas de uma árvore filogenética.

Os nós correspondem aos pontos que são ligados por diversos ramos (linhas). Dentre os nós, existem os mais externos, chamados de nós terminais, que identificam os indivíduos, genes ou proteínas incluídos na análise filogenética; e os nós internos, que representam uma inferência evolutiva do ancestral comum. A ligação entre os nós terminais e os nós internos ocorre por meio dos ramos terminais; e a ligação entre diferentes nós internos é feita por meio de ramos internos (Braun *et al.*, 2014).

Como as análises filogenéticas podem envolver genes, proteínas, indivíduos, espécies, gêneros, famílias, dentre outros, os nós terminais são também denominados OTUs. A disposição das OTUs na filogenia é denominada topologia. A partir da análise da topologia é possível identificar grupos monofiléticos ou clados, que incluem todos os membros descendentes de um único ancestral; grupos parafiléticos, que se originam de um único ancestral, mas não reúnem todos os organismos derivados do ancestral comum; e os grupos polifiléticos, que englobam OTUs com características comuns, mas que são oriundos de dois ou mais ancestrais diferentes (Braun *et al.*, 2014).

Outra estrutura importante nas árvores filogenéticas é a raiz. Esta corresponde à um grupo externo, uma OTU ancestral a todo o grupo que está sob análise, conferindo uma direção temporal à árvore e permitindo observar mudanças que partem das sequências mais antigas (mais próximas à raiz) até as sequências mais recentes (mais distantes). Árvores com ausência de raiz não indicam o ancestral comum, sendo mais comuns falhas na interpretação (Braun *et al.*, 2014).

Após o alinhamento das sequências, um modelo evolutivo deve ser escolhido para descrever matematicamente as mudanças de nucleotídeos ou aminoácidos ao longo do tempo, levando em consideração processos de substituição, inserção, deleção, duplicação, transição (troca de purina por outra purina; ou de pirimidina por outra pirimidina) e transversão (troca de purina por pirimidina ou vice-versa). A distância genética é a medida que expressa a magnitude das alterações entre as sequências, podendo indicar uma relação evolutiva mais ou menos próxima. Em regra, a distância pode ser inferida a partir do tamanho dos ramos, mas uma escala de distância que especifique o número de mudanças ao longo do ramo confere maior confiabilidade a esta análise (Braun *et al.*, 2014).

O modelo de substituição mais simples para sequências de nucleotídeos foi proposto por Jukes e Cantor em 1969 (JC69). Este modelo assume que as mudanças entre nucleotídeos podem ocorrer com a mesma probabilidade, atribuindo 25% de chance para cada nucleotídeo ocorrer na sequência. Em 1980, as primeiras publicações a respeito de genomas mitocondriais levaram à necessidade de criar um modelo que considerasse que transições eram mais comuns que transversões. Assim, foi proposto um novo modelo por Kimura, o K80 ou K2P, que no ano seguinte foi aperfeiçoado com a adição de um novo parâmetro, passando a ser chamado K3P. Posteriormente, Felsenstein criou o modelo F81, baseado no fato de que alguns genomas apresentam maior quantidade de guaninas e citosinas comparado com timinas e adeninas. Diante da maior frequência de algumas bases, seria esperado que algumas substituições fossem mais frequentes que outras, o que foi considerado para computar diferentes proporções de cada nucleotídeo neste modelo. O modelo de Hasegawa, Kishino e Yano (HKY85) combina os modelos K2P e F81, ao supor que a frequência das bases é variável e permitir que transições e transversões ocorram com taxas diferentes. Um modelo mais complexo, resultado da combinação dos anteriores, é o modelo geral reversível no tempo (GTR - *general time reversible*), que incorpora seis parâmetros para taxas de substituição e quatro parâmetros para frequência das bases. Extensões incorporadas ao GTR são a adição de proporção de sítios invariáveis nos alinhamentos (+I) e da taxa de variação entre os sítios (+G) (Luo *et al.*, 2010; Braun *et al.*, 2014).

Dentre os diversos modelos existentes de substituição de aminoácidos, as matrizes PAM consideram que estes são substituídos aleatoriamente e de forma independente, conforme uma probabilidade inerente ao próprio aminoácido. Já as matrizes BLOSUM são estabelecidas considerando a observação de regiões de proteínas distantes evolutivamente, atribuindo porcentagens mínimas de identidade aos blocos de aminoácidos utilizados para construir o alinhamento. Por exemplo, a matriz BLOSUM62 agrupa sequências com pelo menos 62% de

identidade. Outro modelo é a matriz WAG, que aperfeiçoou modelos de substituição anteriores (Braun *et al.*, 2014).

Os modelos evolutivos propostos diferem principalmente quanto ao número de parâmetros aplicados para explicar as mudanças de nucleotídeos ou aminoácidos, podendo ser mais ou menos complexos. A maior complexidade nem sempre resultará no melhor resultado, uma vez que o maior número de parâmetros pode aumentar a probabilidade de erros. Assim, é necessário avaliar cada caso, o que pode ser feito com o auxílio de ferramentas que selecionam o modelo de substituição que melhor se adapta a cada alinhamento, como Modeltest e Jmodeltest (Braun *et al.*, 2014). Um dos parâmetros mais utilizados pelos preditores de modelos evolutivos é o critério de informação Bayesiano (BIC), que combina simplicidade computacional e desempenho eficaz para escolha do modelo mais adequado. A seleção do modelo parte do pressuposto de que os candidatos são igualmente prováveis *a priori*, sendo comparados a partir da aplicação de probabilidades baseadas na estatística Bayesiana, para escolha do modelo mais provável *a posteriori* dentro daquele conjunto de dados. O melhor modelo será aquele com o menor valor de BIC (Neath e Cavanaugh, 2012).

A próxima etapa é a escolha de métodos quantitativos ou qualitativos para construir a árvore. Os métodos quantitativos utilizam as diferenças entre as sequências alinhadas para calcular uma árvore final, a partir de matrizes de distância, e compreendem o método de agrupamento par a par usando médias aritméticas não ponderadas (UPGMA - *unweighted pair-group method using arithmetic averages*) e a aproximação de vizinhos (*neighbor joining*). A principal vantagem dos métodos quantitativos é a eficiência computacional, uma vez que apresentam apenas uma filogenia como resultado final, sendo este o principal ponto negativo. Os métodos qualitativos, representados pela máxima parcimônia, máxima verossimilhança e inferência Bayesiana, inferem a filogenia conforme critérios específicos. Estes métodos envolvem algoritmos que atribuem valores a todas as filogenias que são construídas e as ordenam segundo os fatores estabelecidos a fim de identificar a filogenia que melhor explica a relação evolutiva entre as OTUs. A principal vantagem é a robustez destes métodos, entretanto o custo computacional elevado é um desafio importante (Braun *et al.*, 2014).

O UPGMA é considerado o método mais simples para reconstrução filogenética. Este consiste em uma matriz que mede a distância evolutiva computada para par, e considera que todas as linhagens evoluem a uma taxa de substituição de nucleotídeos ou proteínas constante. Caso os dados não satisfaçam esse pressuposto, árvores com topologias incorretas podem ser geradas (Braun *et al.*, 2014).

O método de aproximação de vizinhos se baseia no princípio de evolução mínima, e consiste na identificação do par que apresenta a menor distância a partir de uma matriz. Este par será unido por um nó, incorporado na árvore e utilizado para calcular uma nova matriz de distância corrigida, que será aplicada para calcular a distância dos pares vizinhos. Este processo vai se repetindo a cada rodada, até que seja dada a árvore que agrupa ao máximo as sequências (Braun *et al.*, 2014).

A máxima parcimônia parte do princípio de que a natureza opta pelos caminhos mais simples, de modo que a filogenia com menor número de eventos evolutivos é considerada a mais provável. A probabilidade de ocorrência de cada nucleotídeo em cada sítio do alinhamento é calculada por uma matriz de custo, segundo a qual substituições de nucleotídeos são penalizadas. Os menores custos determinam a topologia da árvore. Por simplificar demasiadamente os processos evolutivos, esta metodologia vem perdendo espaço, especialmente quando se trata de linhagens com taxas evolutivas variáveis ou muito rápidas (Braun *et al.*, 2014).

A máxima verossimilhança busca a árvore que melhor explica os dados do alinhamento dentre todas as filogenias construídas, a partir de modelos estatísticos complexos. Sua aplicação exige a construção de uma filogenia inicial, geralmente por métodos quantitativos. A esta árvore inicial é aplicado um modelo evolutivo, que irá atribuir parâmetros que auxiliarão no cálculo da probabilidade de uma sequência ocupar determinada posição na topologia da árvore. Para cada sítio do alinhamento será calculada a verossimilhança, e ao final os valores de cada sítio serão multiplicados para encontrar a verossimilhança de cada uma das árvores geradas, que serão então comparadas. Assim, a árvore que apresentar o maior valor de verossimilhança é a que melhor reflete a história evolutiva, considerada a de máxima verossimilhança. Este método estima as relações evolutivas com maior confiabilidade, mas exige alto poder computacional e maior tempo para análise (Braun *et al.*, 2014).

A inferência Bayesiana, por sua vez, engloba a máxima verossimilhança, mas inclui informações disponíveis *a priori*, como a taxa evolutiva, distribuição geográfica, dados de dinâmica populacional, dentre outros. Esses parâmetros são incorporados na forma de modelos de probabilidade, e são levados em consideração para o cálculo da árvore mais provável. O resultado final considera também probabilidades posteriores, como a de uma topologia estar correta de acordo com os dados do alinhamento, e é calculado para cada filogenia proposta a partir de parâmetros especificados pelo usuário. Dado o alto custo computacional e a lentidão das análises, a estatística Bayesiana utiliza amostragens para estimar os valores esperados de

cada parâmetro. Este método calcula a filogenia de maneira robusta, e pode ser utilizado para análises temporais e de dinâmica populacional sob uma ótica evolutiva (Braun *et al.*, 2014).

Por fim, deve ser feita a validação da inferência filogenética, para garantir maior precisão. Esta validação ocorre por meio de algoritmos que produzem amostragens distintas a partir do mesmo conjunto de dados, partindo do princípio de que quanto mais a mesma filogenia se repetir maior será a probabilidade de ser adequada. Dentre os diferentes algoritmos de validação, o método de *bootstrap* é um dos mais utilizados. O algoritmo reconstrói a filogenia a partir do alinhamento dado e, posteriormente, diversas replicatas são construídas. Os sítios do alinhamento são aleatoriamente amostrados e agrupados para formar novos alinhamentos com o mesmo número de sítios do alinhamento original, conforme o número de vezes e o método de construção especificado pelo usuário. Ao final, o algoritmo irá atribuir a confiabilidade da topologia de cada clado conforme o número de vezes que esta se repetiu dentre as replicatas. A maioria dos trabalhos científicos utiliza 1000 replicatas para validar a filogenia, e confiabilidades acima de 70% dadas pelo método de *bootstrap* são consideradas sustentadas. Ressalta-se que na inferência Bayesiana a validação estatística é inerente ao processo, sendo considerados relevantes valores de confiabilidade acima de 90%. (Braun *et al.*, 2014).

Todos os métodos para inferência filogenética apresentam vantagens e desvantagens, reunidas na tabela 2. A escolha adequada será aquela avaliada caso a caso e em conformidade com uma justificativa biológica.

Tabela 2 - Principais características dos métodos de construção de árvores filogenéticas

(Adaptada de Braun *et al.*, 2014)

Tipo	Método de construção	Princípio	Vantagens	Desvantagens
Métodos quantitativos	UPGMA	Agrupamento sequencialmente as OTUs com a menor distância evolutiva entre si	Menor exigência de poder computacional	Maior probabilidade de erros
	Aproximação de vizinhos	Busca a árvore com menor soma total de ramos		
Métodos qualitativos	Máxima parcimônia	Busca a filogenia com menor número de eventos evolutivos	Menor exigência de poder computacional dentre os métodos qualitativos	Simplifica de forma excessiva os processos evolutivos
	Máxima verossimilhança	Busca a árvore com maior valor de verossimilhança dentre as filogenias construídas	Robustez	Mais lenta que inferência Bayesiana; Alta demanda computacional
	Estatística Bayesiana	Realiza a amostragem de filogenias e busca a mais provável	Robustez; Mais rápida que máxima verossimilhança	Parâmetros devem ser especificados; Alta demanda computacional

3.2.3.1 Análises de filogenia para *Senecavirus A*

A maior acessibilidade das tecnologias de sequenciamento ao longo dos anos resultou em uma maior abundância de estudos sobre a epidemiologia molecular de *Senecavirus A*. A filogenia é uma das ferramentas utilizadas nestes trabalhos, permitindo a análise da relação evolutiva entre amostras virais isoladas em diferentes momentos e regiões geográficas.

Vannucci e colaboradores (2015) sequenciaram duas amostras de SVA isoladas de porcas afetadas em Minas Gerais e Goiás e avaliaram a proximidade dessas com linhagens dos Estados Unidos e Canadá. Para isso, construíram duas árvores filogenéticas de máxima verossimilhança, com 1000 *bootstraps*. A primeira, a partir da região VP1, e a segunda baseada em parte da proteína 3Dpol e região 3'UTR completa. Nos dois casos, foi demonstrado que as amostras brasileiras possuíam maior proximidade com a amostra canadense do que com as oriundas dos EUA. A identidade com amostras disponíveis no *GenBank* foi de 94,2 a 96,5%.

Também utilizando a região da proteína VP1, Laguardia-Nascimento e colaboradores (2016) construíram árvore filogenética de máxima verossimilhança, utilizando Kimura 2 como modelo evolutivo e 1000 *bootstraps*. Nesta árvore foram analisadas sequências parciais de SVA isoladas no Rio Grande do Sul, Goiás, Minas Gerais e Santa Catarina, que se agruparam no

mesmo clado, indicando uma origem genética comum. Estas amostras se mostraram próximas às isoladas nos EUA no ano de 2015, e mais relacionadas à sequência canadense de 2011, abordada no estudo de Vannucci e colaboradores mencionado anteriormente, do que às sequências norte-americanas dos anos de 1988 a 1997. A taxa de mutação calculada para as sequências de SVA brasileiras foi de 4×10^{-3} /sítio de substituição/ano, considerada consistente com outros picornavírus.

Joshi e colaboradores (2016b) executaram reconstruções filogenéticas baseadas na região VP1 e também em sequências completas do genoma, as quais revelaram nítida separação entre amostras históricas (anteriores a 2002) e contemporâneas (2007 a 2015). De fato, a identidade genética entre isolados históricos e contemporâneos variou de 86 a 88% em relação ao gene VP1 e 93 a 94% quanto ao genoma completo; enquanto a identidade compartilhada dentre todas as amostras contemporâneas foi superior, de 96 a 99%. Novamente a filogenia indicou proximidade entre as amostras brasileiras e as amostras estadunidenses de 2015. As árvores foram construídas segundo o método de aproximação de vizinhos, com 1000 replicatas.

A separação entre amostras históricas e contemporâneas também foi observada por Leme e colaboradores (2017), com agrupamento das sequências em três cladogramas temporais distintos a partir da comparação da região VP1. O primeiro clado inclui amostras identificadas inicialmente, como SVV-001; o segundo clado inclui sequências dos EUA identificadas entre 1988 e 1997; e o terceiro clado reúne os vírus de diferentes países detectados entre os anos de 2001 e 2016. A árvore construída pelo método de aproximação de vizinhos, com 1000 replicatas, evidenciou ainda o agrupamento das sequências contemporâneas conforme a localização geográfica. Esta característica também foi observada no trabalho de Chen *et al.* (2017) e de Joshi e colaboradores (2020). Neste último, a árvore filogenética foi construída com 117 sequências completas do genoma de SVA, utilizando o método de máxima verossimilhança, o modelo GTR e 500 replicatas. Segundo os autores, o agrupamento por região geográfica encontrado sugere que a maioria das amostras virais evoluiu de forma independente dentro da população suína de cada país afetado, mas a relação filogenética observada entre algumas linhagens chinesas e uma colombiana com amostras dos EUA indica a possibilidade de transferência do vírus entre áreas geográficas distintas. Os autores afirmam também que as mutações identificadas nas sequências de nucleotídeos das amostras históricas, em comparação com as contemporâneas, indicam evolução viral contínua (Joshi *et al.*, 2020).

Vieira e colaboradores (2022) descrevem a terceira onda de surtos de doença vesicular por *Senecavirus A* no Brasil, utilizando árvore de máxima verossimilhança construída a partir do gene VP1 e baseada no modelo Kimura 2, com 1000 *bootstraps*. Nesta, foi evidenciado o

agrupamento temporal das sequências brasileiras, com as amostras de 2018 e 2020 formando um *cluster* separado das amostras de 2015 e 2016. A utilização de sequência do vírus da febre aftosa como raiz da árvore (*outgroup*) resultou em ramos excessivamente curtos, realçando a proximidade genética entre as sequências de SVA (identidade de 93 a 95,9% entre amostras brasileiras e amostras contemporâneas de outros países; e de 96,1 a 98,6% entre as amostras brasileiras). A partir das análises realizadas, os pesquisadores afirmam que as ondas de surtos de SVA estão ocorrendo no Brasil em intervalos cada vez mais curtos, com redução do intervalo de 3 para 1,5 anos, o que buscam associar à evolução molecular do vírus, que pode estar associada à maior patogenicidade das amostras contemporâneas; à duração da imunidade, que ainda não está bem estabelecida; e/ou à taxa anual de reposição de reprodutores (40%), que culmina na substituição de todas as porcas do rebanho a cada 2 a 2,5 anos.

O trabalho de Wu e colaboradores (2022) avaliou a evolução e dinâmica espaço-temporal global de *Senecavirus A*, a partir de 249 sequências completas utilizadas para construção de árvore filogenética por inferência Bayesiana. De acordo com os autores, as sequências podem ser divididas em dois clados principais: o clado I, formado pelas amostras anteriores ao ano de 2007, e o clado II, posterior a esse período. A taxa evolutiva calculada foi de $3,3522 \times 10^{-3}$ /sítio de substituição/ano, com 27 locais da poliproteínas sob forte seleção positiva. As linhagens chinesas foram associadas à maior diversidade genética em comparação com outros países. Os autores também afirmam que as análises demonstraram que o Brasil seria a fonte mais provável de surtos a partir de 2015. Entretanto, trabalhos recentes a respeito dos primeiros relatos de doença vesicular por SVA no México (dezembro/2021) e no Chile (abril/2022) demonstraram que os vírus detectados nestes países são mais próximos geneticamente de linhagens estadunidenses isoladas em 2017 e 2020, respectivamente (Bennet *et al.*, 2022; Navarro-Lopez *et al.*, 2023). Esta percepção está alinhada com os resultados da inferência Bayesiana apresentada por Gao *et al.* (2022), que demonstraram a propagação histórica do vírus a partir dos Estados Unidos e identificaram este país como a principal fonte de transmissão viral. Neste trabalho, a taxa evolutiva calculada foi de $3,93 \times 10^{-3}$ /sítio de substituição/ano, mais semelhante à calculada para as amostras brasileiras por Laguardia-Nascimento e colaboradores, em 2016.

3.2.4 Predições *in silico*

A constante evolução das ferramentas de bioinformática culminou em um novo campo: a imunoinformática. Esta inclui o uso de algoritmos e métodos computacionais para estudar interações patógeno-hospedeiro, identificar fatores de virulência e prever estruturas

imunológicas que possam ser utilizadas no desenvolvimento de vacinas, testes diagnósticos e fármacos (Bahrami *et al.*, 2019).

A estratégia de produção de vacinas a partir de alvos identificados por ferramentas de bioinformática é conhecida como vacinologia reversa. A grande vantagem desta metodologia é limitar os testes *in vitro* e *in vivo* a uma menor quantidade de moléculas, garantindo economia e agilidade no avanço das pesquisas para controle de doenças. A primeira etapa do processo é a predição de proteínas por ferramentas computacionais, seguida da expressão e purificação destas, imunização dos animais e avaliação da imunidade induzida (Bruno *et al.*, 2015). Vários estudos na área de medicina veterinária vêm aplicando esta metodologia, para diferentes patógenos, como *Brucella melitensis* (Vishnu *et al.*, 2015), *Cryptosporidium parvum* (Tomazic *et al.*, 2018) e vírus da febre aftosa (Qadeer *et al.*, 2021).

Uma abordagem de destaque na vacinologia reversa, especialmente para patógenos virais, é a construção de proteínas multi-epítipo que reúnam determinantes antigênicos de células B e T para estimular a resposta imune humoral e celular. A metodologia, revisada em detalhes por Shawan *et al.* (2023), é iniciada com a obtenção da sequência de proteína do patógeno de interesse, em plataformas como NCBI (www.ncbi.nlm.nih.gov/protein/) ou Uniprot (www.uniprot.org). Em seguida, a sequência é submetida a programas como ABCPred (Saha e Raghava, 2006) ou BepiPred (Larsen *et al.*, 2006), para busca de epítopos de linfócitos B; e a outros como NetMHCpan e NetMHCIIpan (Reynisson *et al.*, 2020) para predição de ligantes de MHC de classe I e II, respectivamente. Selecionados os epítopos, conforme os parâmetros destacados a seguir, estes podem ser conectados pelos chamados *linkers*, pequenas sequências ligantes que atuam na expressão, dobramento, estabilização e apresentação antigênica da proteína construída (Chen *et al.*, 2013). Sequências de adjuvantes proteicos podem também ser incluídas para maior estimulação do sistema imune (Shawan *et al.*, 2023). Um parâmetro utilizado para avaliar a afinidade de ligação de peptídeos ao MHC é a concentração inibitória média (IC50). Os programas fornecem resultados em unidades de IC50nM, sendo os peptídeos com valores de $IC_{50} < 50nM$ considerados de alta afinidade (Zhao e Sher, 2018).

Posteriormente, são avaliados aspectos como antigenicidade, alergenicidade, toxicidade, solubilidade e parâmetros físico-químicos. Exemplos de programas utilizados para estes fins são VaxiJen (Doytchinova e Flower, 2007), AllerTOP (Dimitrov *et al.*, 2014), ToxinPred (Gupta *et al.*, 2013), Protein-Sol (Hebditch *et al.*, 2017) e ProtParam (Gasteiger *et al.*, 2005). O ProtParam fornece informações sobre o número de aminoácidos, peso molecular, ponto isoelétrico (indica o comportamento básico ou ácido da proteína), composição de

aminoácidos, carga dos resíduos, composição e fórmula atômica, coeficiente de extinção (indica a absorção de luz por determinada proteína em um comprimento de onda específico, sendo útil para purificações em espectrofotômetro), meia-vida estimada (previsão do tempo necessário para metade da quantidade de uma proteína em uma célula desaparecer após sua síntese), índice de instabilidade (estimativa da estabilidade da proteína em um tubo de ensaio, onde proteínas com $II < 40$ são consideradas estáveis), índice alifático (volume relativo ocupado por cadeias laterais alifáticas, com valores maiores associados à maior estabilidade da proteína em faixas diferentes de temperatura), e grande média de hidropaticidade (GRAVY –valores negativos indicam a hidrofiliabilidade e valores positivos sugerem hidrofobicidade da proteína) (Gasteiger *et al.*, 2005).

Para avaliação da estrutura da proteína, programas como PSIPRED (McGuffin *et al.*, 2000), ColabFold (Mirdita *et al.*, 2022), AlphaFold (Jumper *et al.*, 2021), RaptorX (Xu *et al.*, 2021), e PROCHECK (Laskowski *et al.*, 1993) estão disponíveis. Uma forma de medir a confiabilidade da predição da estrutura tridimensional da proteína é o teste de diferença de distância local prevista (pLDDT), que produz uma estimativa de confiança por resíduo em uma escala de 0 a 100. Regiões modeladas com pLDDT > 90 são consideradas de alta precisão; de 70 a 90 são bem modeladas; de 50 a 70 são de baixa confiabilidade e aquelas com pLDDT < 50 possuem aparência de fita e sugerem desordem estrutural (Tunyasuvunakool *et al.*, 2021). A estrutura predita pode também ser avaliada a partir de *Ramachandran plot*, gerado pelo PROCHECK. Para cada resíduo em uma proteína, são possíveis dois ângulos de torção, denominados ϕ (phi) e ψ (psi). Um gráfico de *Ramachandran* (figura 19) mostra os ângulos phi-psi, o que permite inferir a estrutura da proteína, uma vez que existem combinações de ângulos típicos para cada estrutura: no quadrante A estão as combinações de folha- β ; no quadrante B estão as combinações de α -hélice esquerda; no quadrante C estão as combinações alfa-hélice direita e no quadrante D estão ângulos de torção estereoquimicamente inviáveis, sob risco de colisão entre átomos. As cores mais quentes representam as regiões mais favoráveis de valores phi-psi, enquanto a escala de cores mais claras indica gradativamente resíduos em regiões menos favoráveis, que podem comprometer a funcionalidade da proteína. Idealmente, 90% dos resíduos de uma proteína devem estar localizados nas regiões mais favoráveis (Birbeck College, [s.d.]; Laskowski *et al.*, 1993). Caso não seja atingida a porcentagem mínima considerada ideal, a estrutura tridimensional pode ser refinada a partir de programas como GalaxyRefine (Heo *et al.*, 2013), mediante protocolos baseados em dinâmica molecular ou atribuição de pontuações para diferenciar conformações nativas de não-nativas (Adiyaman e McGuffin, 2019).

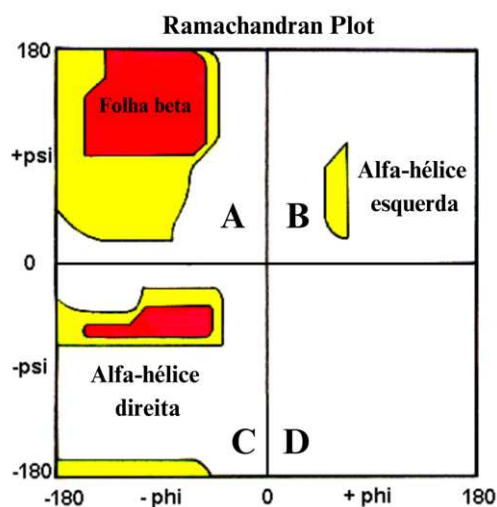


Figura 19 - Adaptada de Birbeck College, [s.d.]. Representação das estruturas secundárias de proteínas por quadrante do *Ramachandran plot*. Em vermelho estão os resíduos localizados em regiões mais favoráveis; em amarelo as regiões menos permissivas; e em branco são as regiões consideradas proibidas devido ao risco de colisão entre átomos.

Além disso, podem ser realizadas análises mais complexas, como ancoragem (*docking*) molecular (Torchala *et al.*, 2013; Yan *et al.*, 2017), simulação de resposta imunológica (Rapin *et al.*, 2010), inclusive quanto à indução da produção de interferon (Dhanda *et al.*, 2013), e a clonagem *in silico* (Grote *et al.*, 2005; SnapGene, [s.d.]).

Alguns desafios enfrentados ao trabalhar com vacinologia reversa são o desenvolvimento laboratorial complexo e oneroso após a predição da vacina, o que demanda diversas etapas, recursos financeiros e infraestrutura (Plotkin *et al.*, 2017); a impossibilidade de prever antígenos não proteicos, como polissacarídeos, que atuam como componentes ativos incrementando a imunogenicidade (Heinson *et al.*, 2015); e falhas na indução de imunidade *in vivo*, por fatores associados à estrutura ou expressão da proteína (Bahrami *et al.*, 2019), ou por limitações das predições *in silico* para modelos animais, uma vez que a maioria dos programas disponíveis foram desenvolvidos para utilização com dados humanos (Bilić *et al.*, 2018).

A predição *in silico* de epítopos também vem sendo utilizada no desenvolvimento de métodos diagnósticos sorológicos. Ferramentas de bioinformática voltadas para linfócitos B, como ABCPred (Saha e Raghava, 2006) e ElliPro (Ponomarenko *et al.*, 2008) estão disponíveis, respectivamente, para identificar epítopos lineares, conectados sequencialmente na cadeia polipeptídica; e epítopos conformacionais destes, que não estão dispostos de forma sequencial, mas se aproximam a partir de dobramentos da proteína (Zhou *et al.*, 2022). Os epítopos de

linfócitos B selecionados podem ser sintetizados para uso de forma isolada, ou associados na construção de proteína multi-epítipo, semelhante ao discutido para vacinologia reversa. A vantagem que se pretende obter com o uso destes epítipos em testes diagnósticos é a melhoria da sensibilidade, a partir da escolha racional e associação dos melhores alvos disponíveis; e da especificidade, pela eliminação de determinantes antigênicos inespecíficos encontrados nas proteínas nativas. Os inconvenientes envolvem principalmente limitações inerentes aos algoritmos de predição, cujos resultados nem sempre coincidem com o observado nos testes *in vitro*; e o custo associado à obtenção dos peptídeos preditos, que ao longo dos anos vem se tornando mais acessível (Vale *et al.*, 2021; Zhou *et al.*, 2022).

3.2.4.1 Predições *in silico* para *Senecavirus A*

Diversos grupos de pesquisa vêm aplicando ferramentas de bioinformática para prever e/ou analisar determinantes antigênicos de *Senecavirus A*, com o objetivo de utilizá-los no desenvolvimento de vacinas e de testes diagnósticos. Conforme abordado anteriormente, as proteínas VP1, VP2 e VP3 são externas ao capsídeo viral, razão pela qual frequentemente são alvo dos estudos voltados para a identificação de epítipos.

O primeiro trabalho a respeito de epítipos de SVA foi desenvolvido por Fan e colaboradores (2020), que não utilizaram abordagens *in silico*, mas sim anticorpos monoclonais e técnicas como *Western blot* e ELISA indireto para mapear epítipos de células B das proteínas VP1 e VP2. Como resultados, seis epítipos lineares foram identificados, sendo um da proteína VP1 (²¹GELAAP²⁶) e cinco da proteína VP2 (¹²DRVITQT¹⁸, ⁷¹WTKAVK⁷⁶, ⁹⁸GGAFTA¹⁰³, ¹⁵⁰KSLQELN¹⁵⁶ e ²⁴⁸YKEGAT²⁵³). Chen *et al.* (2021) utilizaram uma metodologia semelhante, com foco na proteína VP3, encontrando o epítipo linear de linfócito B ¹⁹²GWFSLHKLTK²⁰¹. Ma e colaboradores (2022) sintetizaram peptídeos sobrepostos da proteína VP2 e utilizaram iELISA para mapear o epítipo ²⁷¹GLRNRFTTGTDEEQ²⁸⁴, que foi então empregado como antígeno para desenvolvimento de iELISA para fins diagnósticos. Nos três trabalhos mencionados, a utilização da bioinformática se restringiu a análises quanto à conservação e à estrutura dos epítipos.

Rieger e colaboradores (2023) utilizaram ferramentas de bioinformática para alinhar as sequências brasileiras disponíveis nos bancos de dados, com a finalidade de verificar se estas possuíam os epítipos encontrados no trabalho de Fan *et al.* (2020) mencionado. Além disso, foi gerado modelo 3D da proteína VP2, com indicação da localização dos epítipos.

Utilizando uma abordagem *in silico*, Zhang e colaboradores (2023) identificaram quatro epítipos de células B na proteína VP1 (posições de aminoácidos: 7–26, 48–74, 92–109 e 129–

144) e cinco na proteína VP2 (posições: 38–57, 145–160, 154–172, 193–208, 249–284). A sequência utilizada foi a linhagem chinesa CH-FuJ-2017 (*Genbank* n. MH490944). Análises de hidrofobicidade, estrutura secundária, flexibilidade, valor antigênico e probabilidade de superfície foram conduzidas no software DNASTar (DNASTar Inc., EUA). Epítomos de células B identificados no estudo e dois epítomos universais de células T foram utilizados para construção de três genes multi-epítopo. Sequências GGC ou KK foram utilizados como *linkers* entre os epítomos de células B e os de células T; enquanto os fragmentos de epítomos de células B foram conectados entre si pelo ligante GGSSGG. Após etapas de validação, os genes foram expressos em plasmídeo e as proteínas obtidas foram utilizadas para a imunização de leitões. Dentre as formulações, a denominada rP2 foi considerada mais eficaz, induzindo 80% de proteção contra o desafio com amostra homóloga.

Ru *et al.* (2023) avaliaram a imunogenicidade e atividade de neutralização de quatro epítomos de linfócitos B, oriundos da proteína VP2, preditos por ferramentas de bioinformática (⁴¹TKSDPPSSSTDQPTTT⁵⁶, ¹⁴⁵PDGKAKSLQELNEEQW¹⁶⁰, ¹⁶¹VEMSDDYRTGKNMPF¹⁷⁵ e ²⁶⁷PYFNGLRNRFTTGT²⁸⁰). As análises foram baseadas na mesma linhagem chinesa utilizada no trabalho de Zhang *et al.* (2023). Além das predições *in silico*, foi utilizada uma abordagem *Pepscan*. Os quatro peptídeos foram sinteticamente produzidos para imunização de cobaias e demonstraram capacidade de induzir resposta humoral específica. Testes de neutralização *in vitro* foram conduzidos após a imunização e obtenção do antissoro, com resultados promissores associados ao segundo peptídeo, correspondente aos resíduos 145-160.

4 MATERIAL E MÉTODOS

4.1 Amostra viral

A amostra viral utilizada foi isolada durante o surto de *Senecavirus A* ocorrido no Brasil no ano de 2015. Esta amostra foi gentilmente cedida pelo Professor Roberto Guedes, do Laboratório de Patologia Molecular da Escola de Veterinária da UFMG. Após a multiplicação do vírus em monocamada de células da linhagem NCI-H1299 (ATCC® CRL-5803TM), estoques virais foram obtidos e congelados a -80°C. A amostra isolada por cultivo celular foi denominada LPVA4. Na ocasião do congelamento, o título viral calculado pela técnica de *Reed and Muench* (1983) foi de $2 \times 10^{9,42}$ TCID₅₀/mL. Ressalta-se que as etapas referentes à obtenção dos estoques virais foram realizadas por Brenda Rocha, durante a execução de seu projeto de mestrado. A metodologia detalhada pode ser consultada na dissertação publicada (Rocha, 2022).

4.2 Preparo da amostra para sequenciamento

Para o envio da amostra para sequenciamento, inicialmente foi realizada a extração de RNA com o *QIAamp Viral RNA Mini Kit* (*Qiagen Inc.*, Alemanha) e a análise do rendimento pelo fluorômetro *Qubit* (*InvitrogenTM*, EUA). A técnica de RT-PCR foi executada para confirmar a presença do RNA viral no inóculo utilizado. Para tal, foi utilizado o kit *SuperScriptTM III One-Step RT-PCR* (*InvitrogenTM*, EUA), os primers 5'-GAAGGTGCGCAGCAGGATGAT-3' (*forward*) e 5'-GGTCGTAGACAAAGCTGGAAGCCTGG-3' (*reverse*), para amplificação do gene da proteína VP1 (Knowles *et al.*, 2006); e as seguintes ciclagens: 45°C por 30 minutos, 94°C por 2 minutos, 40 ciclos de 94°C por 30 segundos, 60°C por 45 segundos e 68°C por 45 segundos; e um ciclo de 68°C por 5 minutos (protocolo desenvolvido no laboratório). O material extraído foi enviado para a empresa Neopropecta (*Neopropecta Microbiome Technologies*, Santa Catarina, Brasil), que procedeu à síntese do DNA complementar (cDNA) necessário para o sequenciamento, como prestação de serviço.

4.3 Sequenciamento da amostra e análise de qualidade

A amostra LPVA4 foi sequenciada pela empresa Neopropecta, via prestação de serviços, utilizando a plataforma de sequenciamento de nova geração MiSeq (*Illumina Inc.*, EUA). A qualidade do sequenciamento foi avaliada pela ferramenta FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/).

4.4 Mapeamento e montagem do genoma

A montagem do genoma foi realizada em parceria com o Laboratório de Bioinformática Integrativa (LBI), coordenado pela Professora Flávia Aburjaile. Três programas foram utilizados para comparação e escolha do melhor resultado: *Bowtie2* (www.bowtie-bio.sourceforge.net/bowtie2/), *Burrows-Wheeler Aligner* (BWA - bio-bwa.sourceforge.net) e RNASTAR (code.google.com/p/rna-star/). Em todos os casos a montagem foi realizada após remoção inicial de conteúdos de adaptadores pelo programa Trimmomatic (www.usadellab.org/cms/index.php?page=trimmomatic) e a partir do mapeamento contra a sequência referência SVV-001 (acesso *Genbank*: NC_011349).

4.5 Alinhamento e análise filogenética

Sequências de *Senecavirus A* disponíveis no NCBI *Virus* (www.ncbi.nlm.nih.gov/labs/virus/vssi/#/) foram selecionadas para alinhamento com a sequência LPVA4 (data de acesso para obtenção das sequências: 30/10/2024). Como critério de escolha, foram priorizadas sequências com tamanho superior a 7.000 pares de bases, visando aquelas com tamanho próximo do esperado para a espécie (~7.3kb). Informações sobre as sequências utilizadas, como o número de acesso *Genbank*, estão disponíveis no Apêndice I. No total, 280 sequências, incluindo a LPVA4, foram alinhadas utilizando o MAFFT (<https://mafft.cbrc.jp/alignment/software/>). Quanto à análise filogenética, o método utilizado foi a máxima verossimilhança, com 1000 *bootstraps*. O modelo evolutivo utilizado, GTR + G + I, foi escolhido a partir da análise de modelos evolutivos do MEGA 11.0 (*Molecular Evolutionary Genetics Analysis* - www.megasoftware.net/). A construção da árvore foi feita no software IQTree (www.iqtree.org), e o FigTree 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) foi utilizado para visualização e edição. A árvore foi enraizada com a linhagem referência SVV-001 (acesso *Genbank*: **131**), semelhante ao trabalho de Joshi *et al.* (2016b) e Buckley *et al.* (2021). Uma matriz de distância foi gerada pelo MEGA 11.0 para avaliar a distância genética entre as 14 sequências brasileiras depositadas no NCBI *Virus*.

4.6 Predição de epítomos e avaliação da conservação

Para as etapas de predição, foi utilizada a sequência SVA/BRA/MG1/2015 (acesso *Genbank*: KR063107), também isolada em Minas Gerais, Brasil, no mesmo ano que a sequência LPVA4 (2015). As análises foram realizadas com proteínas estruturais e não estruturais de *Senecavirus A*. A predição de ligantes de MHC de classes I e II foi realizada a partir das ferramentas NetMHCpan e NetMHCIIpan, acessadas pela plataforma *TepiTool* (<http://tools.iedb.org/tepitool/>), do *Immune Epitope Database & Tools* (IEDB). Para MHC I, foram preditos epítomos com 9 aminoácidos; enquanto para MHC II o comprimento estabelecido foi de 15 aminoácidos. Foram selecionados os epítomos com $IC_{50} \leq 25nM$, visando maior afinidade de ligação. A ferramenta MHC I *Immunogenicity* (<http://tools.immuneepitope.org/immunogenicity/>), do IEDB, foi também utilizada para filtrar os epítomos de MHC I com maior potencial imunogênico. Para predição de epítomos lineares de linfócitos B, com 16 aminoácidos de comprimento, o programa *ABCPred* (<https://webs.iitd.edu.in/raghava/abcpred/>) foi utilizado. Os epítomos de linfócitos B, T CD4+ e T CD8+ foram concatenados para remoção de redundâncias. Além disso, os epítomos lineares

de linfócitos B preditos pelo *ABCPred* com *score* maior ou igual a 0.90 foram selecionados para posterior avaliação do potencial para uso em métodos sorológicos. A conservação dos epítomos selecionados foi avaliada a partir do alinhamento das proteínas, feito pela ferramenta *AliView* (<https://ormbunkar.se/aliview/>), utilizando as sequências elencadas nos Apêndices II e III.

4.7 Construção de proteínas multi-epítomo

Os epítomos preditos que preencheram os requisitos mencionados no item anterior foram conectados com *linkers* para obtenção da sequência da proteína multi-epítomo para vacina (M-E1). O ligante AAY foi utilizado para os epítomos de linfócitos T citotóxicos (ligantes de MHC I); enquanto GPGPG foi utilizado para conectar os epítomos de linfócitos T auxiliares (ligantes de MHC II), conforme trabalho de Vilela Rodrigues e colaboradores (2022). Seguindo o exemplo de estudos para outros patógenos virais que acometem suínos, foram testados três adjuvantes, conectados no início da sequência construída pelo ligante EAAAK: cadeia B da enterotoxina termolábil de *Escherichia coli* (Sun *et al.*, 2013), GPX1 - glutationa peroxidase-1 suína (Du *et al.*, 2017) e proteína de choque térmico GP96, também conhecida como endoplasmina (Chen *et al.*, 2013). Os epítomos das proteínas do capsídeo com *score* maior ou igual a 0.90, preditos pelo programa *ABCPred*, contendo 16 aminoácidos, foram utilizados para construção de outra proteína multi-epítomo (M-E2), voltada para fins diagnósticos.

4.8 Análises de antigenicidade, alergenicidade e toxicidade

As duas proteínas multi-epítomo construídas foram submetidas a análises de antigenicidade, alergenicidade e toxicidade. Estas análises foram realizadas, respectivamente, pelos programas: *VaxiJen 2.0* (www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html), *AllerTOP 2.0* (www.ddg-pharmfac.net/AllerTOP/ToxinPred) e *ToxinPred* (<http://webs.iitd.edu.in/raghava/toxinpred/protein.php>) (Vilela Rodrigues *et al.*, 2022).

4.9 Avaliação de parâmetros físico-químicos

Parâmetros como massa molecular, ponto isoelétrico (pI), coeficiente de extinção, índice alifático, grande média de hidropaticidade (GRAVY), meia-vida estimada para organismos modelo (*E. coli*, levedura e células de mamíferos) e índice de instabilidade das duas proteínas multi-epítomo foram analisados utilizando o *ProtParam*

(<https://web.expasy.org/protparam/>). A solubilidade foi calculada pelo programa *Protein-Sol* (<https://protein-sol.manchester.ac.uk>).

4.10 Análises de homologia

A existência de homologia entre as proteínas construídas e o proteoma de suínos domésticos (*Sus scrofa domesticus*) foi avaliada pela ferramenta *Basic Local Alignment Search Tool* – *Protein* (BLASTp) do servidor NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>), para prevenir reações de autoimunidade ou interferências no diagnóstico (Vilela Rodrigues *et al.*, 2022). A homologia foi verificada também em relação ao proteoma de outros vírus causadores de doença vesicular em suínos (vírus da febre aftosa, estomatite vesicular, doença vesicular dos suínos e exantema vesicular dos suínos), a fim de avaliar riscos quanto à especificidade do diagnóstico.

4.11 Análises estruturais

O programa PSIPRED 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>) foi utilizado para determinar a estrutura secundária das proteínas multi-epítopo; enquanto o *AlphaFold2* (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/v1.2.0/AlphaFold2.ipyn>) foi utilizado para prever a estrutura terciária. O *Ramachandran plot* foi gerado pelo programa PROCHECK, acessado através da ferramenta SAVES (<https://saves.mbi.ucla.edu>). Por fim, o *GalaxyRefine* (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) foi utilizado para refinar as estruturas obtidas, com avaliação das estruturas finais novamente pelo *Ramachandran plot* (Vilela Rodrigues *et al.*, 2022).

4.12 Predição de epítomos conformacionais de células B

As estruturas tridimensionais refinadas das duas proteínas multi-epítopo construídas foram submetidas ao programa *ElliPro* (<http://tools.iedb.org/ellipro/>), com os parâmetros padrão, para busca de epítomos conformacionais de linfócitos B, indispensáveis na estimulação do sistema imune (Vilela Rodrigues *et al.*, 2022).


5 RESULTADOS E DISCUSSÃO

5.1 Preparo da amostra para sequenciamento

O rendimento do extraído da amostra LPVA4 foi satisfatório, de 2.048ng/ml. O resultado do RT-PCR foi positivo.

5.2 Sequenciamento da amostra e análise de qualidade

O sequenciamento de nova geração realizado pela empresa Neopropecta resultou em dois arquivos contendo as sequências brutas do genoma sequenciado, um referente à fita *forward* (R1) e o outro à fita *reverse* (R2), ambos no formato *fastq*. A análise destes arquivos pelo programa FastQC retornou os resultados apresentados nas figuras 20 a 32.

A  **Basic Statistics**

Measure	Value
Filename	230301110085-1-1-1_S457_L001_R1_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1274812
Sequences flagged as poor quality	0
Sequence length	305
%GC	41

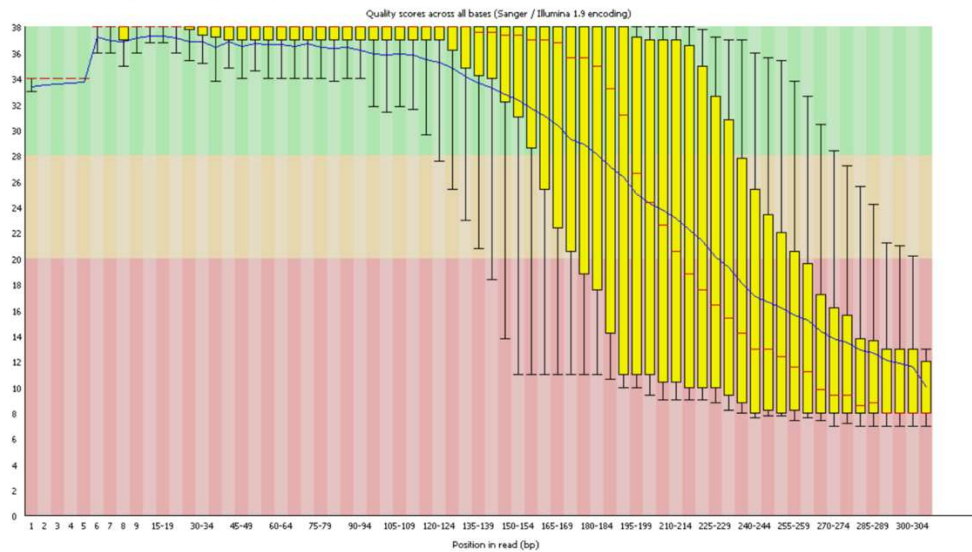
B  **Basic Statistics**

Measure	Value
Filename	230301110085-1-1-1_S457_L001_R2_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1274812
Sequences flagged as poor quality	0
Sequence length	205
%GC	43

Figura 20 - Estatísticas básicas obtidas pelo sequenciamento da amostra LPVA4: fita *forward* (A) e fita *reverse* (B).

As estatísticas básicas mostradas na figura 20 confirmam o uso da plataforma *Illumina* para sequenciamento, com comprimentos de sequência fixos em 305 pares de bases, no caso da fita *forward*, e 205 pares de base na fita *reverse*. O conteúdo GC encontrado foi de 41 e 43%, enquanto no genoma de referência SVV-001 este foi reportado como 51,6% (Hales *et al.*, 2008). A diferença observada pode ser justificada pelo conteúdo celular associado ao vírus sequenciado.

A Per base sequence quality



B

Per base sequence quality

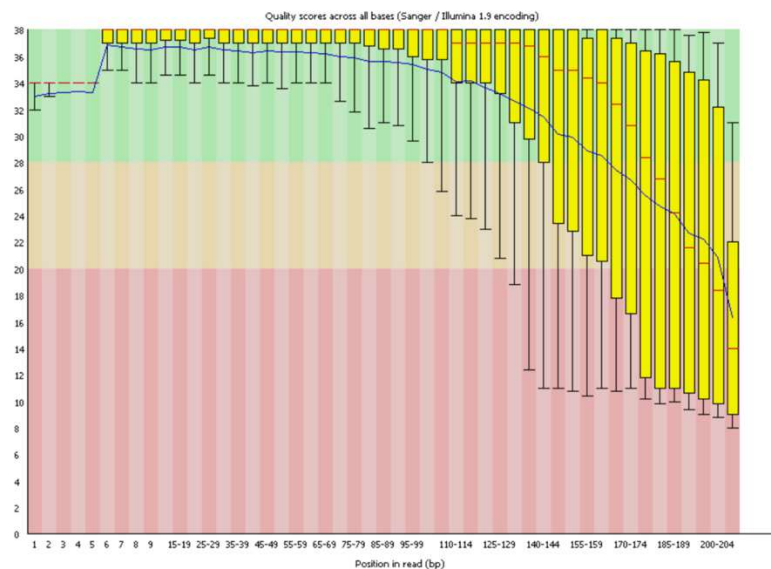


Figura 21 - Qualidade da seqüência por base, nas fitas *forward* (A) e *reverse* (B).

Ao analisar a figura 21, é possível notar uma perda de qualidade do sequenciamento, com a mediana (linha vermelha) de algumas bases apresentando valor inferior à 20, o que ensejou um alerta de erro (símbolo X). No caso da fita *reverse* (figura 21-B), essa perda de qualidade foi mais restrita ao fim do sequenciamento, o que inicialmente estaria dentro do esperado para longas execuções, com perspectiva de melhoria após cortes de qualidade e remoção do conteúdo de adaptadores. Entretanto, ao analisar o resultado conjuntamente com a fita *forward* (figura 21-A), que apresentou queda de qualidade de forma mais precoce, em cerca de metade das bases, entende-se que outro fator pode estar relacionado, como uma

contaminação por adaptadores (figura 30) ou por vestígios do cultivo celular utilizado para obtenção da amostra viral (Babraham Bioinformatics, [s.d.], a).

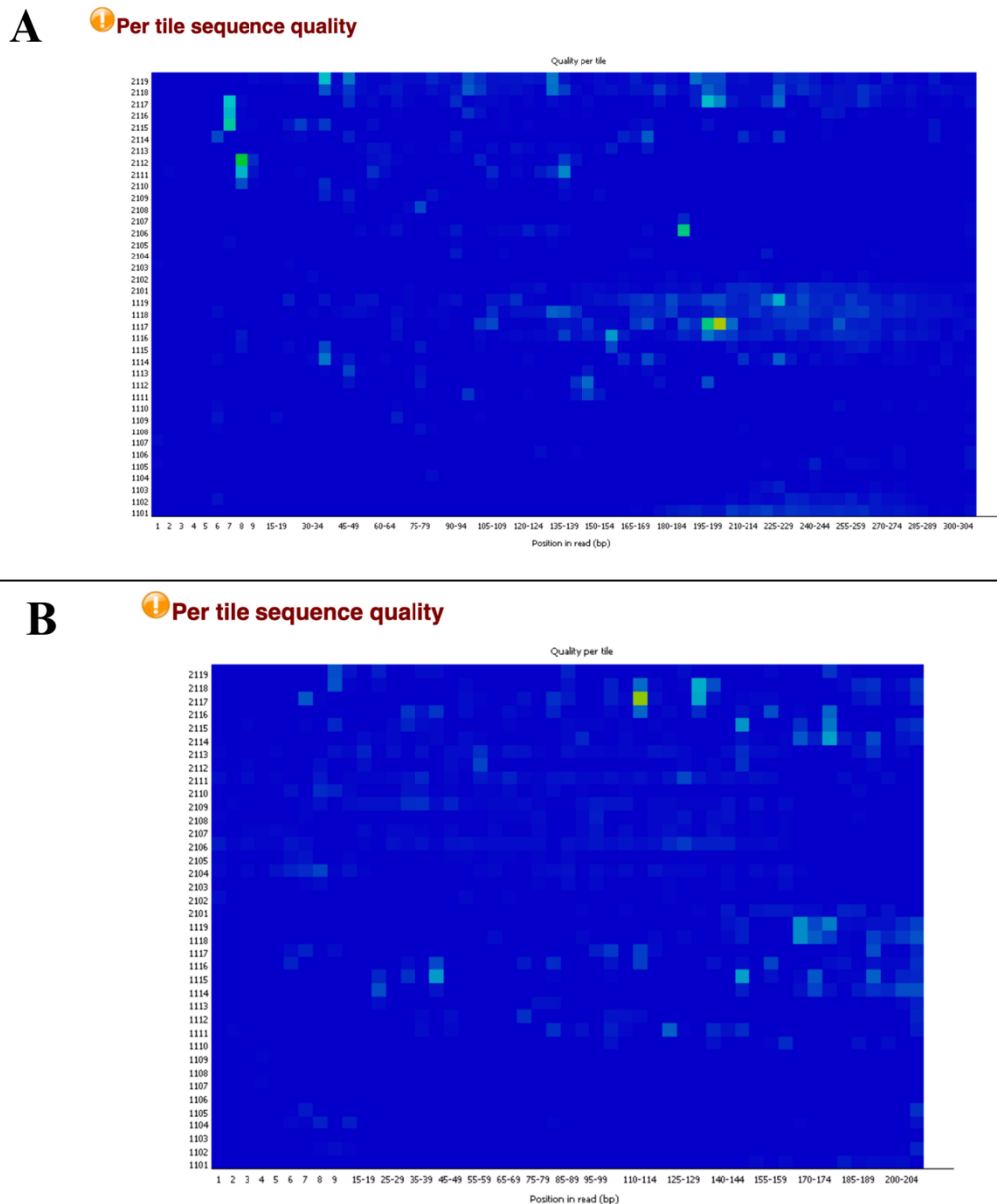


Figura 22 - Qualidade da sequência por bloco, nas fitas *forward* (A) e *reverse* (B).

O predomínio de tons frios observado na figura 22 indica que, na análise da qualidade da sequência por blocos, a maioria destes apresentou qualidade satisfatória. Entretanto, tanto na fita *forward* (figura 22-A) quanto *reverse* (figura 22-B) foi emitida advertência (símbolo !), o que ocorre quando algum bloco apresenta pontuação média de *Phred* (eixo y) mais que duas vezes inferior à média dos demais blocos, corroborando com a perda de qualidade evidenciada na figura 21 (Babraham Bioinformatics, [s.d.], a).

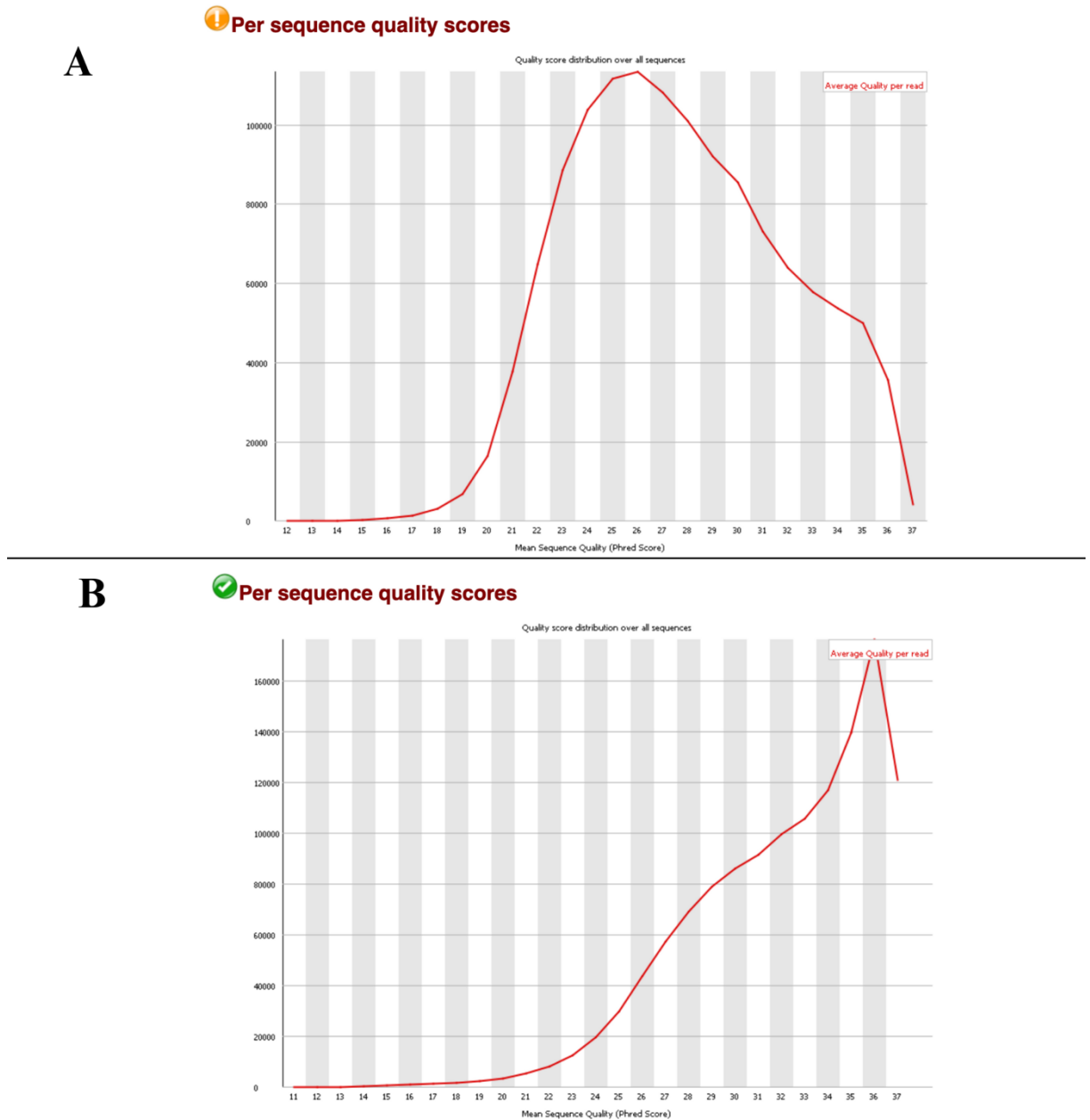


Figura 23 - Scores de qualidade por sequência, nas fitas *forward* (A) e *reverse* (B).

O gráfico apresentado na figura 23-B está coerente com o esperado, com a qualidade média observada com maior frequência equivalente à pontuação *phred* de 37. Por outro lado, na figura 23-A, relacionada à fita *forward*, esta pontuação é de 26, inferior ao mínimo estipulado para que não seja emitida uma advertência (*phred* de 27). Neste contexto, a taxa de erro observada na fita *forward* foi discretamente superior a 0,2%, condizente com o previsto para corridas longas (Babraham Bioinformatics, [s.d.], a).

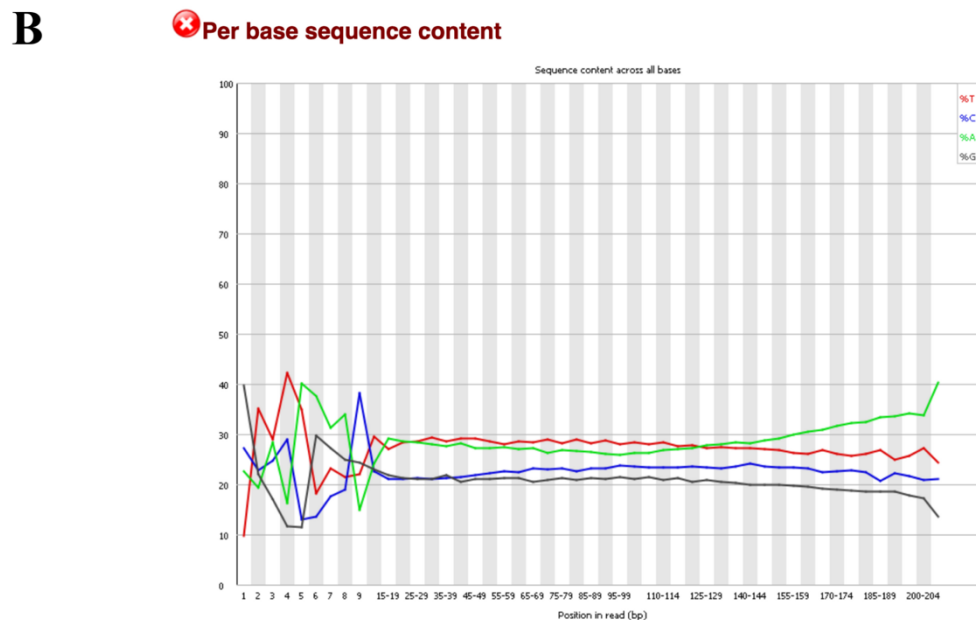
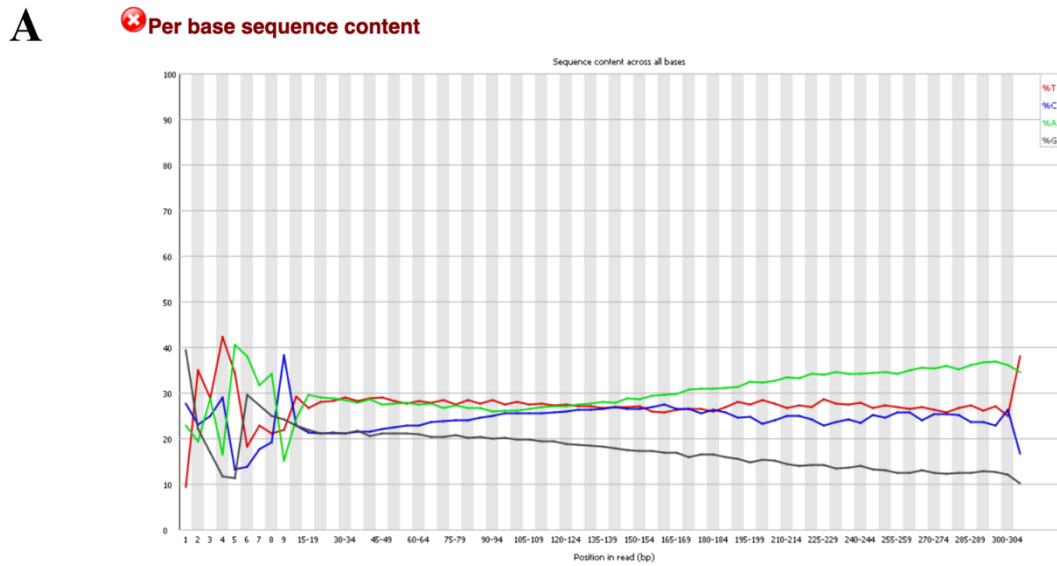


Figura 24 - Conteúdo da sequência por base, nas fitas *forward* (A) e *reverse* (B).

O conteúdo da sequência por base, mostrado na figura 24, demonstra certa desproporção entre as porcentagens de adenina (A), citosina (C), timina (T) e guanina (G) no genoma sequenciado, quando o esperado seria a adequada proporção entre A-T e G-C, com maior paralelismo entre as linhas, indicando a ausência de desequilíbrios marcantes. O padrão observado no início da leitura não desperta preocupação, tendo em vista a utilização de transposase como adaptador, o que pode enviesar as posições de início da leitura. O final também pode ser desconsiderado, por se tratar provavelmente de ruídos do equipamento. Entretanto, quando avaliadas as porcentagens de cada base, nota-se diferença superior a 20%

em algumas posições ao decorrer da leitura, o que justifica o aviso de erro observado e indica possível contaminação, semelhante às outras análises mencionadas (Babraham Bioinformatics, [s.d.], a).

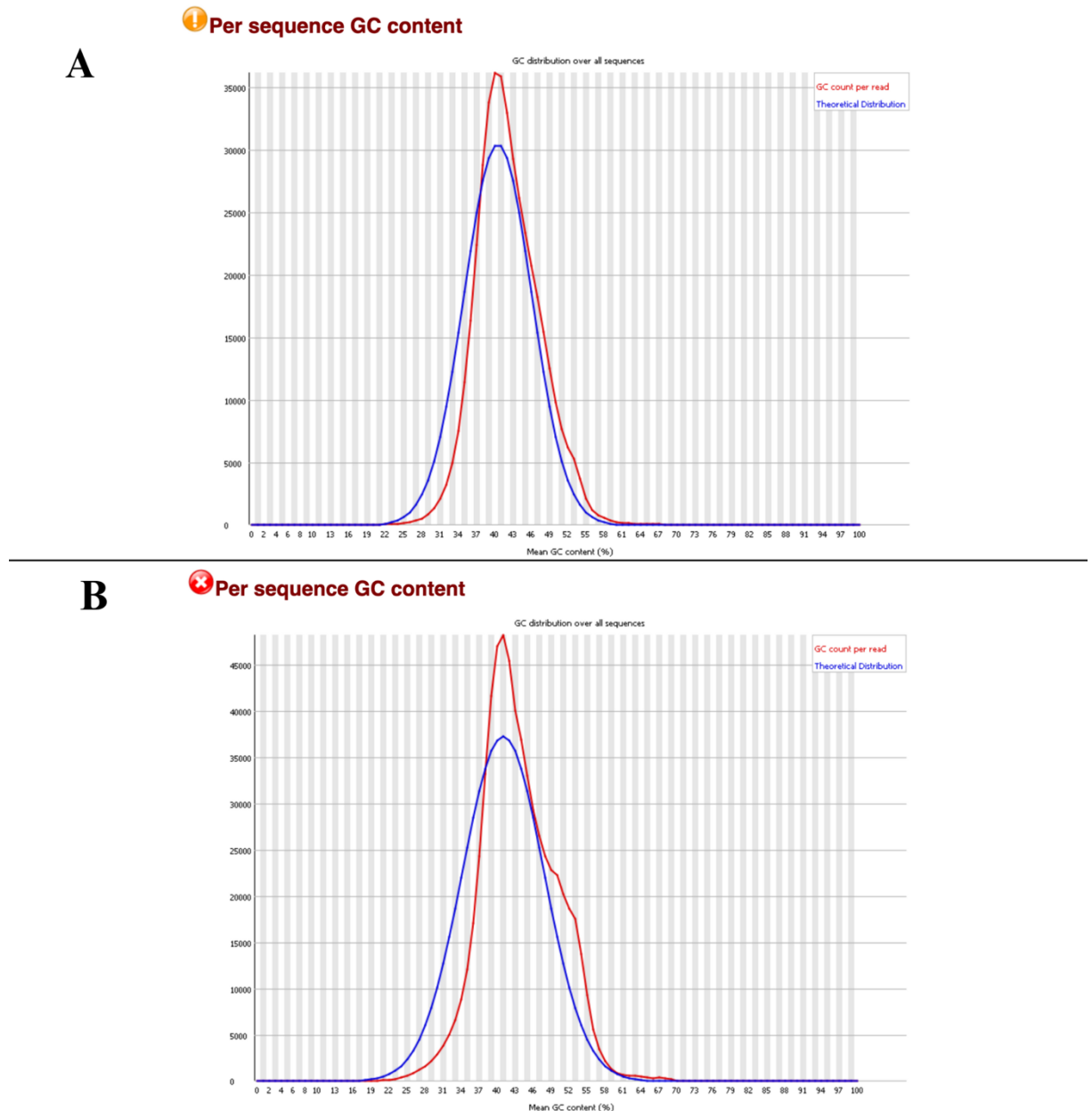
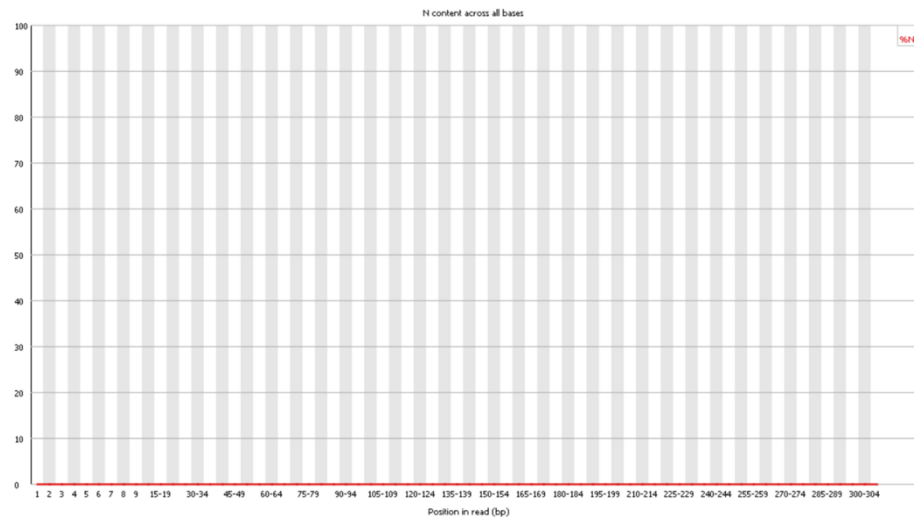


Figura 25 - Conteúdo GC por sequência, nas fitas *forward* (A) e *reverse* (B).

O gráfico da figura 25 representa o conteúdo GC por sequência, sendo esperada distribuição normal. A advertência observada no gráfico referente à fita *forward* (figura 25-A), e o aviso de erro relacionado à fita *reverse* (figura 25-B) ocorrem pela soma dos desvios da

distribuição normal ser superior à 15% e 30% das leituras, respectivamente. Os picos agudos constatados estão associados à presença do adaptador, evidenciada na figura 30 (Babraham Bioinformatics, [s.d.], a).

A  **Per base N content**



B  **Per base N content**

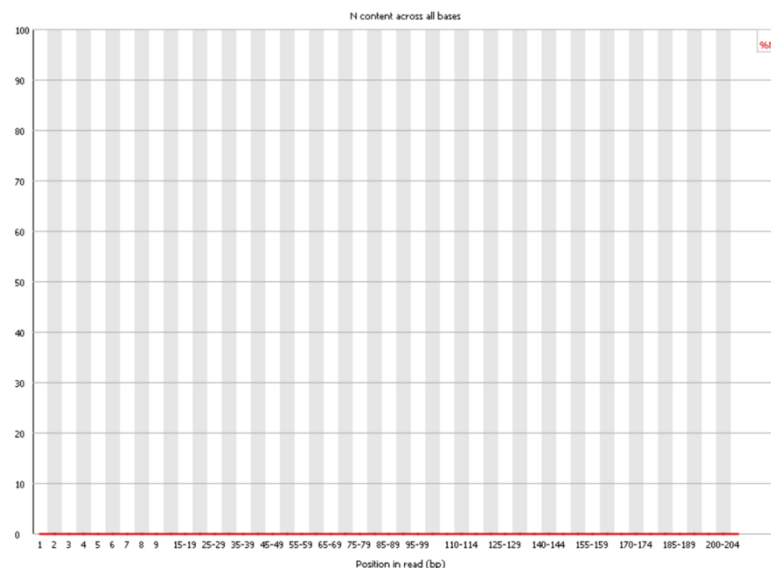


Figura 26 - Conteúdo N por base, nas fitas *forward* (A) e *reverse* (B).

Apesar das falhas reveladas pelos gráficos anteriores, o gráfico apresentado na figura 26 demonstra a ausência de conteúdo N por base, indicando que a plataforma de sequenciamento não teve dificuldade em determinar quais nucleotídeos estavam presentes em quais posições (Babraham Bioinformatics, [s.d.], a). Este resultado favorável reforça a hipótese de que as alterações observadas no sequenciamento até então estão relacionadas ao conteúdo

de adaptadores e ao conteúdo celular, uma vez que a amostra viral utilizada foi obtida após passagens em cultivo de células.

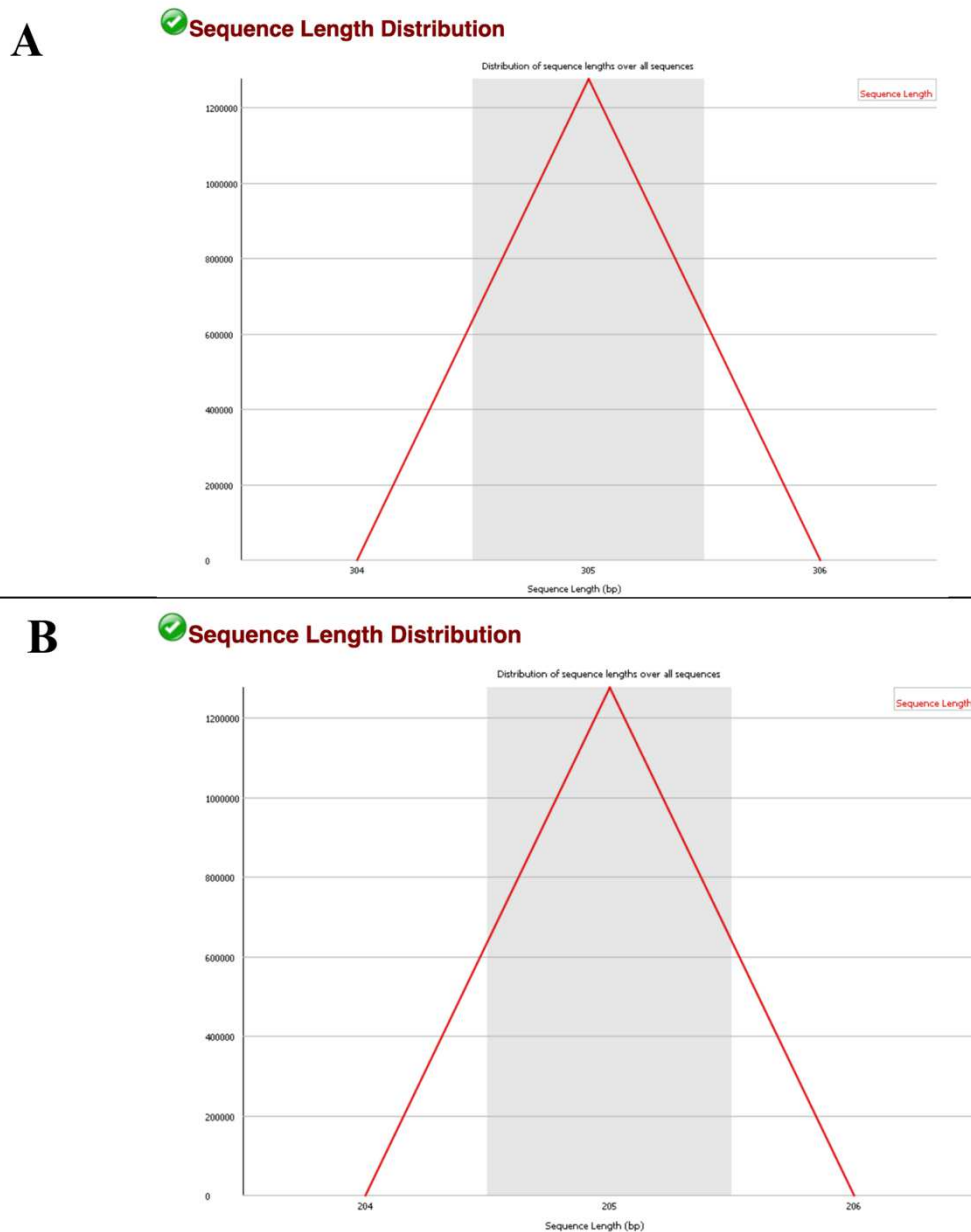


Figura 27 - Distribuição do comprimento de sequência, nas fitas *forward* (A) e *reverse* (B).

A obtenção de fragmentos de comprimento uniforme, evidenciada na figura 27, corrobora com o esperado para utilização da plataforma *Illumina* (Babraham Bioinformatics, [s.d.], a). No caso da fita *forward*, este comprimento foi de 305 pares de base; enquanto na fita *reverse* o comprimento foi de 205 pares de base.

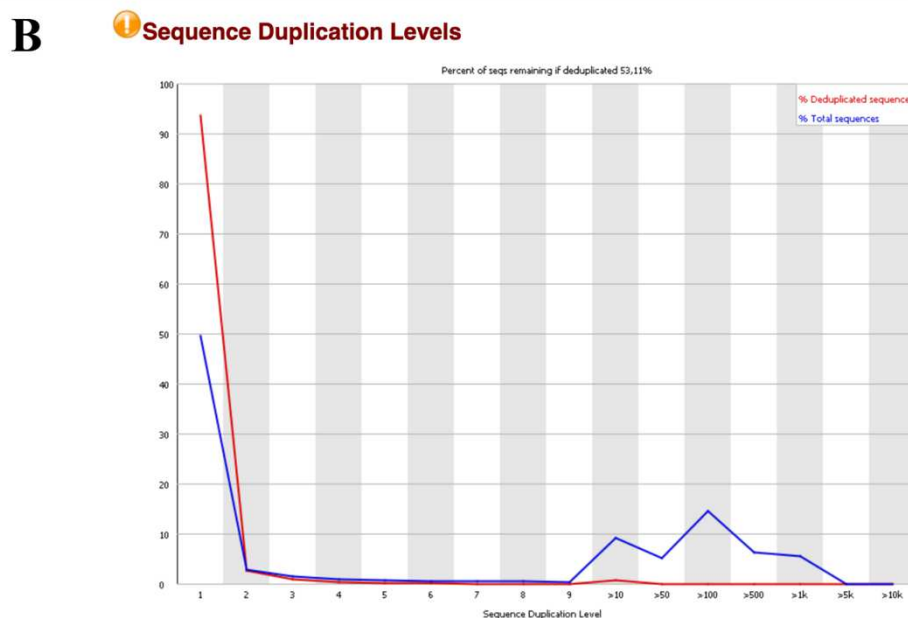
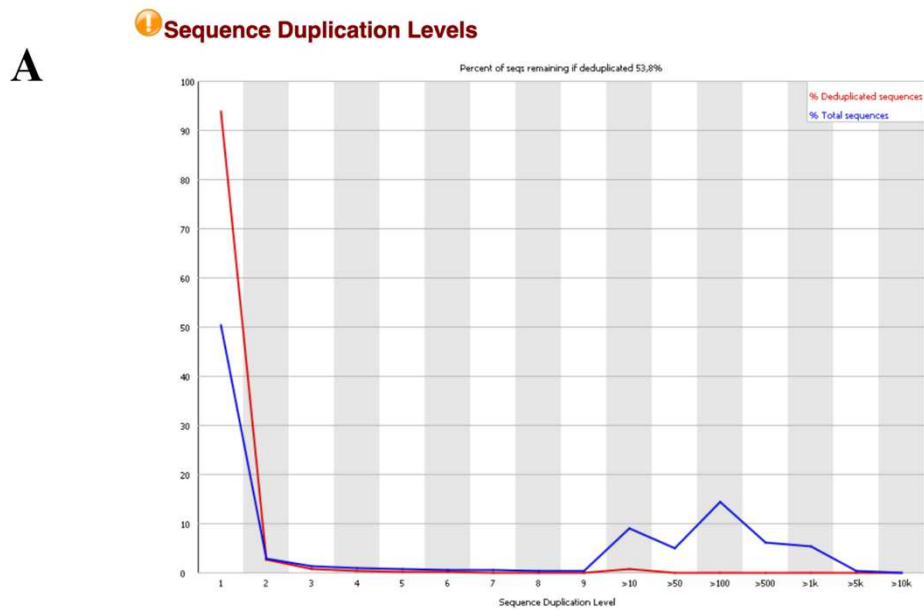


Figura 28 - Nível de duplicação das sequências, nas fitas *forward* (A) e *reverse* (B).

O nível de duplicação de sequências das fitas *forward* e *reverse* foi similar, com observação de picos à direita na linha azul culminando em advertências (figura 28). Estas indicam que as sequências duplicadas representam mais de 20 e menos de 50% do total, o que mais uma vez está correlacionado com a presença de adaptadores e de contaminação celular. Adicionalmente, bibliotecas de RNA-Seq comumente apresentam sequências em níveis diferentes na população inicial, uma vez que para observar transcrições de baixa expressão são sequenciadas a mais transcrições de maior expressão, resultando em grande conjunto de duplicatas (Babraham Bioinformatics, [s.d.], a).

A

Overrepresented sequences

Sequence	Count	Percentage	Possible Source
GTATAGGGGTTAGTCTTGCTATATTATGCTGGTTATAATTTTCATCT	4754	0.37291773218325525	No Hit
CCCCATACCTTCTGCATAATGAATTAACAGAAATAAAGTTGCAAGGAG	3889	0.3050645899160033	No Hit
ACCTTAGCCAAACCATTTACCAAATAAGTATAGGCGATAGAAATTGAA	3380	0.265137133946025	No Hit
GTGTACCACCTGTCGATGGCAGTAAAAGTGGTGGGTTGATCTGTGCT	2772	0.2174438270113554	No Hit
ATATAGTACCGCAAGGGAAGATGAAAAATTATAACCAAGCATAATAG	2272	0.17822235749271265	No Hit
GTGGTACACTGGACGCCTCAATTCTGGACAAAAGCTGAAAAACCTCT	2214	0.1736726670285501	No Hit
CTCTAAATCCCCTTGAAATTAAGTGTAGTCCAAAGGAGAACAGCTCT	2013	0.15790563628205573	No Hit
GATTTAGAGGGTCTGTGGGCAAAATTAAGTTGAACTAAGATTCTATCT	1953	0.15319905993981858	No Hit
CCTAAGAAGCTAAAAGAGCACACCGCTATGTAGCAAAATAGTGGGA	1926	0.15108110058581187	No Hit
AGATATAGTACCGCAAGGGAAGATGAAAAATTATAACCAAGCATAATAG	1878	0.14731583951202218	No Hit
GTTCTAGGTAGCTCGTGGTTTCGGGGGCTTAGCTTTGGCTCTCCTT	1874	0.14700206775587304	No Hit
GGTGATAGCTGGTTGTCGAAGATAGAACTTAGTTCAACTTTAAATTTGC	1791	0.14049130381577832	No Hit
CGTCTATGTAGCAAAATAGTGGGAAGATTATAGGTAGAGGCGACAAACC	1692	0.13272545285108706	No Hit
GTACTATATCTATTGGCCAGGTTCAATTTCTATCGCTATACTTTATT	1680	0.13178413758263965	No Hit
GCTATACACGGCTCGGTAGGTTTTCGCCTTACCTATAAATCTCCCA	1609	0.12621468891099238	No Hit
ATCTATCACCTATAGAAAGCACTAAGTTAGTATAAGTAACATGAAACA	1601	0.1255871453986941	No Hit
ATCTGGACAACAGCTATCACAGGCTCGGTAGGTTTTCGCCTTACC	1587	0.12448894425217208	No Hit
ATATATAGCTGGTTATAATTTTCATCTTTCCCTTGGCGTACTATATCT	1558	0.12221409902009081	No Hit
CCCATAGTAGGCTAAAAGCAGCCCACTAAGAAAGGCTTCAAGCTCA	1553	0.12182188432490439	No Hit
GCCAAGATAGAATCTTAGTTCAACTTTAAATTTGCCACAGAACCTCT	1537	0.12056679730030782	No Hit
GCATAATAGCAAGGACTAACCCCTATACCTTCTGCATAATGAATTAAC	1486	0.11656620740940625	No Hit
GGCTTATGCGGAGGAGAAATGTTTCATGTTACTATACTAACATTAGTTC	1415	0.11099675873775898	No Hit
ACATAGACGGGTGCTCTTTAGCTGTTCTTAGGTAGCTGCTGGTTT	1402	0.10997700053027427	No Hit

B

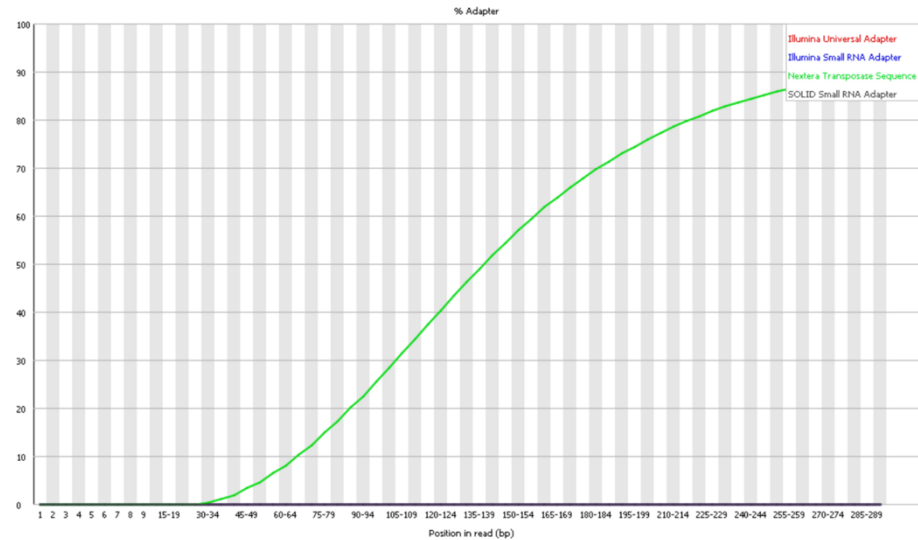
Overrepresented sequences

Sequence	Count	Percentage	Possible Source
GTATAGGGGTTAGTCTTGCTATATTATGCTGGTTATAATTTTCATCT	5090	0.3992745596997832	No Hit
CCCCATACCTTCTGCATAATGAATTAACAGAAATAAAGTTGCAAGGAG	3912	0.306868775138609	No Hit
ACCTTAGCCAAACCATTTACCAAATAAGTATAGGCGATAGAAATTGAA	3212	0.25195872018776105	No Hit
GTGTACCACCTGTCGATGGCAGTAAAAGTGGTGGGTTGATCTGTGCT	2754	0.21603185410868425	No Hit
ATATAGTACCGCAAGGGAAGATGAAAAATTATAACCAAGCATAATAG	2211	0.17343733821143825	No Hit
GTGGTACACTGGACGCCTCAATTCTGGACAAAAGCTGAAAAACCTCT	2147	0.16841699011305197	No Hit
GATTTAGAGGGTCTGTGGGCAAAATTAAGTTGAACTAAGATTCTATCT	2131	0.16716190308845538	No Hit
GTTCTAGGTAGCTCGTGGTTTCGGGGGCTTAGCTTTGGCTCTCCTT	2057	0.16135712559969526	No Hit
CTCTAAATCCCCTTGAAATTAAGTGTAGTCCAAAGGAGAACAGCTCT	2006	0.1573565357087947	No Hit
CCTAAGAAGCTAAAAGAGCACACCGCTATGTAGCAAAATAGTGGGA	1969	0.15445414696441515	No Hit
ATATATAGCTGGTTATAATTTTCATCTTTCCCTTGGCGTACTATATCT	1869	0.1466098530606866	No Hit
GTACTATATCTATTGGCCAGGTTCAATTTCTATCGCTATACTTTATT	1832	0.14370746431630704	No Hit
AGATATAGTACCGCAAGGGAAGATGAAAAATTATAACCAAGCATAATAG	1799	0.1411188473280766	No Hit
CGTCTATGTAGCAAAATAGTGGGAAGATTATAGGTAGAGGCGACAAACC	1741	0.13656915686391405	No Hit
GGTGATAGCTGGTTGTCGAAGATAGAACTTAGTTCAACTTTAAATTTGC	1639	0.12856797708211093	No Hit
GCTATACCAAGGCTCGGTAGGTTTTCGCCTTACCTATAAATCTCCCA	1629	0.12778354769173808	No Hit
ATCTTGGACAACAGCTATCACAGGCTCGGTAGGTTTTCGCCTTACC	1603	0.12574403127676867	No Hit
ATCTATACCCCTATAGAAGCACTAATGTTAGTATAAGTAACATGAAACA	1584	0.12425361543506022	No Hit
GGCTTATGCGGAGGAGAAATGTTTCATGTTACTATACTAACATTAGTTC	1510	0.1184488379463011	No Hit
GTCCAAGATAGAATCTTAGTTCAACTTTAAATTTGCCACAGAACCTCT	1469	0.1152326774457724	No Hit
ACATAGACGGGTGCTCTTTAGTGTCTTAGGTAGCTGCTGGTTT	1448	0.1135853757259894	No Hit
CCCATAGTAGGCTAAAAGCAGCCCACTAAGAAAGCTTCAAGCTCA	1435	0.11256561751850469	No Hit
GCATAATAGCAAGGACTAACCCCTATACCTTCTGCATAATGAATTAAC	1410	0.11060454404257256	No Hit
ACTATATCTATTGCGCCAGGTTCAATTTCTATCGCTATACTTTATTG	1348	0.10574108182226006	No Hit
GGGTAGTCTTGCTATATTATGCTGGTTATAATTTTCATCTTTCCCT	1344	0.10542731006611171	No Hit

Figura 29 - Dados das sequências super-representadas e possível fonte de contaminação (não identificada), nas fitas *forward* (A) e *reverse* (B).

As sequências super-representadas, ou seja, aquelas que representam mais de 0,1% do total, estão elencadas na figura 29. A super-representação pode estar associada à maior importância biológica destas sequências ou à contaminação (Babraham Bioinformatics, [s.d.], a). Tendo em vista que não se tem conhecimento de nenhuma descrição na literatura, até o momento, a respeito de sequências com papel de destaque na biologia de *Senecavirus A*, a segunda hipótese parece ser a mais factível, visto que pode ser inferida pela análise conjunta com os demais resultados.

A Adapter Content



B Adapter Content

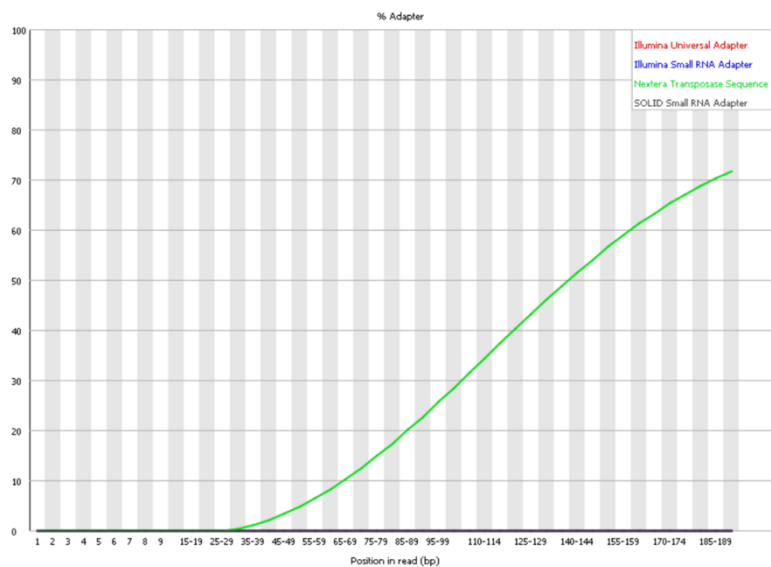


Figura 30 - Conteúdo de adaptadores, nas fitas *forward* (A) e *reverse* (B).

Conforme mencionado, os resultados obtidos pela ferramenta FastQC sugerem a existência de contaminações que resultaram em uma menor qualidade do sequenciamento. A figura 30 confirma a hipótese da presença de adaptadores como uma das fontes de contaminação, sendo identificado o uso de transposase, em consonância com o que foi informado pela empresa responsável pelo sequenciamento. A remoção de adaptadores é uma etapa importante durante o processamento dos dados do sequenciamento, que tende a promover melhoria significativa dos resultados, recomendando-se nova análise por meio do FastQC após este passo.

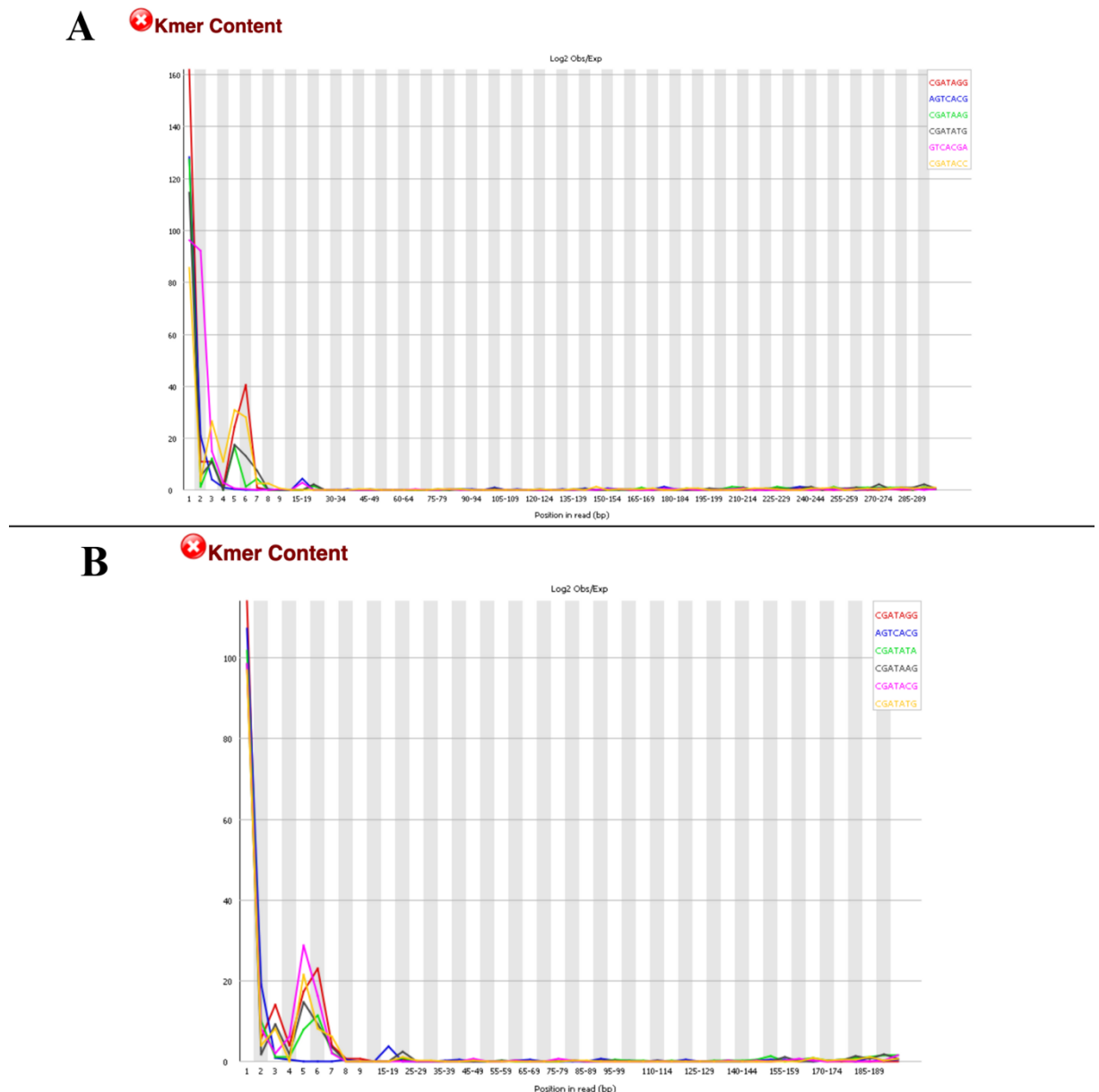


Figura 31 - Conteúdo K-mer, nas fitas *forward* (A) e *reverse* (B). Em cada gráfico estão representadas as seis sequências K-mer mais frequentes.

Corroborando com as demais análises, pode ser observado no sequenciamento conteúdo K-mer (figuras 31 e 32), especialmente na forma de picos acentuados em pontos específicos no início das leituras. Este padrão em regra está associado ao enriquecimento posicional das sequências super-representadas observadas anteriormente, indicativas de contaminação (Babraham Bioinformatics, [s.d.], a).

A

Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CGATAGG	1220	0.0	161.7538	1
AGTCACG	2435	0.0	128.31804	1
CGATAAG	985	0.0	127.49215	1
CGATATG	1370	0.0	114.58007	1
GTCACGA	3385	0.0	96.280464	1
CGATACC	1745	0.0	85.673195	1
CGATATT	1470	0.0	85.42842	1
CGATACG	715	0.0	79.4544	1
TCACGAT	4230	0.0	77.400566	2
AGGGCGG	180	1.8189894E-12	74.75279	5
GATAGGC	655	0.0	73.03803	2
CGATAAT	1755	0.0	68.99988	1
CGATAAA	2135	0.0	67.22236	1
CGATATA	1835	0.0	66.80642	1
GATACGG	390	0.0	65.16655	2
CGATAGC	1450	0.0	59.79989	1
GATATGG	1095	0.0	58.707653	2
GA CTGGG	1020	0.0	55.698162	7
GATAGGT	755	0.0	55.443607	2
CGGTGGG	110	2.767867E-4	54.365665	7

B

Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CGATAGG	1205	0.0	113.95021	1
AGTCACG	2030	0.0	107.34237	1
CGATATA	870	0.0	101.78735	1
CGATAAG	535	0.0	98.57009	1
CGATACG	485	0.0	98.47423	1
CGATATG	965	0.0	96.92228	1
GGGCTCG	115	0.0	86.52174	4
CGATACC	1270	0.0	85.397644	1
CGATAAT	900	0.0	75.17777	1
TAACGGG	40	0.001726229	74.625	4
GTCACGA	3025	0.0	73.35042	2
CGATATT	980	0.0	73.10204	1
CGATATC	805	0.0	72.92547	1
TATGGCG	75	3.910456E-6	66.333336	4
GATACGG	340	0.0	64.382355	2
TCACGAT	3590	0.0	60.974934	3
CGATACT	655	0.0	59.244274	1
GATAGGT	595	0.0	58.529408	2
TCGGAGA	230	0.0	56.239132	8
CACGATA	4145	0.0	52.810616	4

Figura 32 - Dados das seqüências K-mer identificadas nas fitas *forward* (A) e *reverse* (B).

A análise conjunta dos resultados dos gráficos apresentados nas figuras a 20 a 32 evidencia que os desvios observados podem ser justificados pelo conteúdo de adaptadores e conteúdo celular associado à amostra viral sequenciada. A remoção do conteúdo de adaptadores é uma etapa comum após o sequenciamento, tendo sido realizada pelo Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>). Quanto ao conteúdo celular, este é identificado e eliminado a partir do mapeamento da seqüência contra um genoma de referência.

5.3 Montagem do genoma

A montagem do genoma da amostra LPVA4 pelos programas *Bowtie2* e BWA foi semelhante, enquanto a montagem pelo RNASTAR apresentou resultado inferior. A sequência obtida após montagem pelo programa BWA (figura 33), a partir do mapeamento contra a referência SVV-001 (acesso *Genbank*: NC_011349), possui 7.253 pares de base, tamanho compatível com o genoma completo de *Senecavirus A* (Wang *et al.*, 2023).

De maneira geral, o sequenciamento e montagem do genoma foram bem-sucedidos. Entretanto, várias regiões de *gaps* foram observadas na sequência LPVA4 (figura 33). Isto porque, ao fazer o mapeamento contra a referência, o montador não conseguiu identificar e/ou ordenar adequadamente o conteúdo de determinadas regiões do genoma. Este é um inconveniente comum das plataformas que realizam o sequenciamento a partir de leituras curtas, como é o caso da tecnologia *Illumina* utilizada (Quer *et al.*, 2022).

Uma alternativa que pode ser adotada para obter um melhor resultado na montagem da amostra LPVA4 é o mapeamento contra uma sequência mais próxima geneticamente, como a SVA/BRA/MG1/2015 (acesso *Genbank*: KR063107), oriunda do mesmo local (Minas Gerais), ano (2015) e que se alinhou no mesmo clado (figura 35).

Sendo possível realizar novo sequenciamento, uma opção é a amplificação do genoma por primers específicos, com utilização do produto da PCR para o sequenciamento. Desta forma, é possível obter menor quantidade de resíduos celulares, que contaminam a sequência e prejudicam o processamento dos dados, sem que haja prejuízos à boa cobertura do genoma (Quer *et al.*, 2022). Outra possibilidade, caso haja disponibilidade de equipamento, é a realização de sequenciamento de leitura longa por meio de plataforma de terceira geração, como *Nanopore*. Leituras longas favorecem a correta montagem do genoma, uma vez que menos *contigs* serão gerados para serem ordenados, com menor risco de que estes sejam posicionados de maneira incorreta ou resultem na inserção de *gaps* (Quer *et al.*, 2022).

5.4 Alinhamento e análise filogenética

Após o alinhamento das 280 sequências completas de *Senecavirus A* (amostra LPVA4 e 279 sequências elencadas no Apêndice I), foi feita a predição do modelo evolutivo GTR + I + G como o mais adequado para a inferência filogenética, devido ao menor valor BIC apresentado (figura 34) (Neath e Cavanaugh, 2012). A árvore construída a partir deste modelo, pelo método de máxima verossimilhança e 1000 replicatas, está representada na figura 35.

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)
GTR+G+I	567	141612.449	134510.640	-66688.162	0.44	0.89	8.27	0.235	0.246
TN93+G+I	564	141685.552	134621.317	-66746.502	0.44	0.89	9.92	0.235	0.246
HKY+G+I	563	141720.579	134668.869	-66771.278	0.44	0.84	8.28	0.235	0.246
GTR+G	566	141810.234	134720.950	-66794.317	n/a	0.30	8.21	0.235	0.246
K2+G+I	560	141810.379	134796.243	-66837.967	0.44	0.82	9.45	0.250	0.250
T92+G+I	561	141816.506	134789.845	-66833.767	0.44	0.83	8.44	0.241	0.241
HKY+G	562	141907.163	134867.977	-66871.833	n/a	0.29	8.27	0.235	0.246
TN93+G	563	141925.242	134873.532	-66873.610	n/a	0.30	9.17	0.235	0.246
K2+G	559	142009.298	135007.687	-66944.689	n/a	0.29	8.42	0.250	0.250
T92+G	560	142022.640	135008.504	-66944.097	n/a	0.28	8.42	0.241	0.241
GTR+I	566	143228.559	136139.275	-67503.480	0.59	n/a	8.16	0.235	0.246
TN93+I	563	143356.052	136304.342	-67589.015	0.58	n/a	8.42	0.235	0.246
HKY+I	562	143374.228	136335.042	-67605.366	0.59	n/a	8.24	0.235	0.246
K2+I	559	143502.338	136500.726	-67691.209	0.59	n/a	8.35	0.250	0.250

Figura 34 - Comparação entre os diferentes modelos evolutivos, conforme o alinhamento realizado. O modelo GTR + G + I foi considerado o mais adequado, uma vez que seu *Bayesian Information Criterion* (BIC) foi inferior (Neath e Cavanaugh, 2012).

Das 279 sequências obtidas no NCBI *Virus*, 265 informam no próprio banco de dados a data de coleta da amostra, e apenas 24 revelam neste o material biológico utilizado para isolamento, sendo citados cavidade abdominal, tecidos (sem especificação), cérebro, banda coronária, fezes, lesões, pulmão, nasofaringe, tonsilas, linfonodo e baço – individualmente ou em combinação (informações extraídas do NCBI *Virus*, disponíveis no Apêndice I). Isto, associado ao fato de que não são incluídos genomas completos na respectiva página, mas apenas na seção de nucleotídeos, demonstra a necessidade de que a deposição de genomas em bancos de dados seja realizada com maior cuidado, uma vez que informações simples, porém relevantes, poderiam ser incluídas sem dificuldade e enriquecer diversas análises.

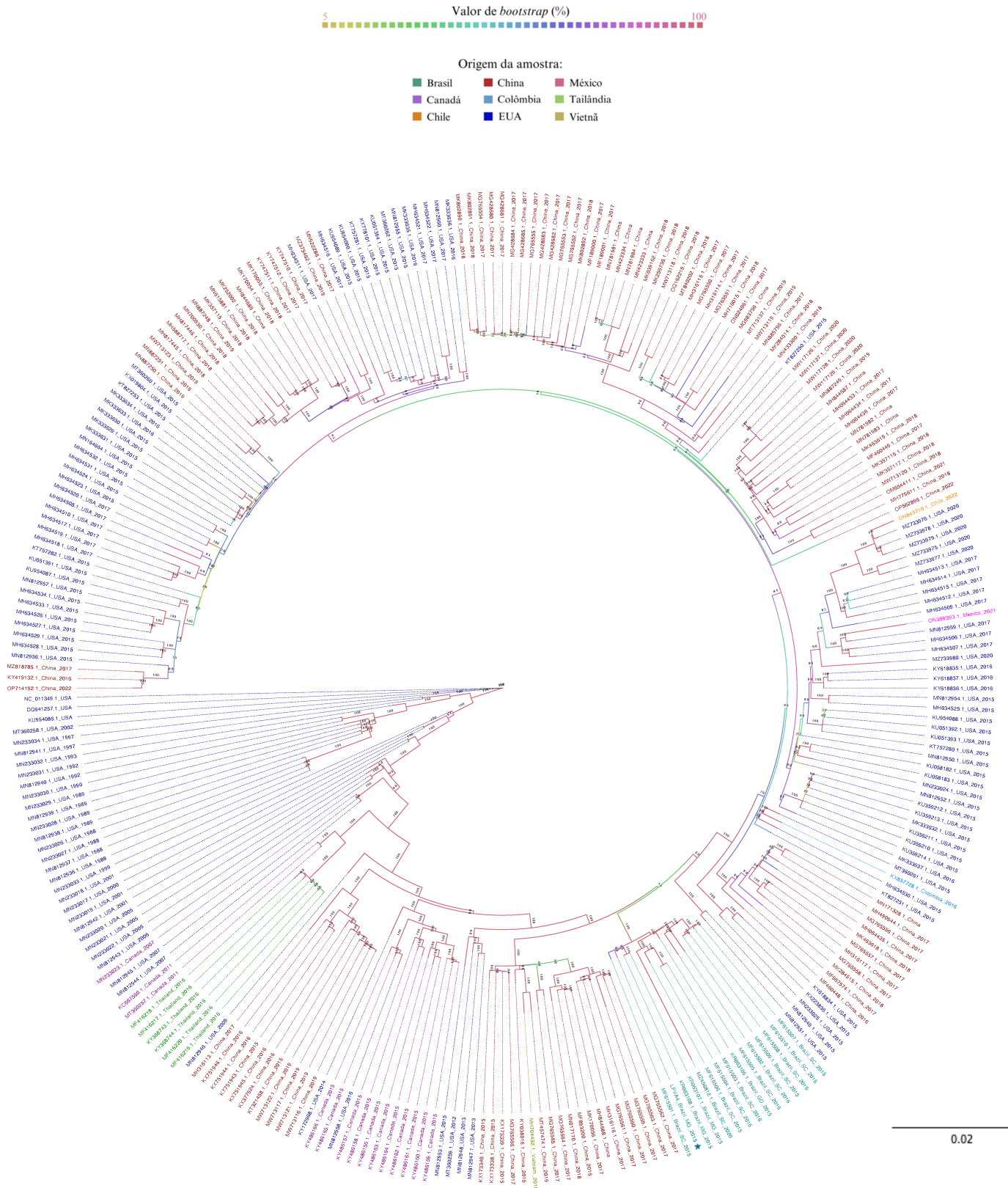


Figura 35 - Árvore de máxima verossimilhança construída a partir do modelo GTR + I + G, com 1000 *bootstraps*, obtida pelo IQTree e editada com o FigTree. Foram utilizadas 280 seqüências, incluindo a LPVA4, destacada com o símbolo estrela. O número de acesso, origem e data de coleta de cada amostra (quando disponível) estão indicados em cada táxon. As cores nos ramos representam os valores de *bootstrap*, com cores mais quentes indicando maior confiabilidade, conforme legenda; enquanto as cores utilizadas em cada táxon indicam o país de origem da amostra.

A árvore construída corrobora com os resultados de filogenia encontrados por outros grupos de pesquisa, com tendência de agrupamento das sequências conforme o ano de isolamento e a região geográfica (Leme *et al.*, 2017; Joshi *et al.*, 2020). A relação filogenética entre linhagens de diferentes países, que levou à formulação da hipótese de transferência viral entre áreas geográficas distintas, por Joshi *et al.* (2020), também foi observada na árvore obtida no presente trabalho.

Assim como observado por Joshi e colaboradores (2016b), as 15 amostras brasileiras se agruparam e se aproximaram das amostras estadunidenses do ano de 2015. A amostra MZ456812.1, isolada em Santa Catarina no ano de 2020, apresentou ramo notadamente maior do que as demais sequências brasileiras, o que levou à análise de matriz de distância. O resultado, exposto na figura 36, demonstra que todas as 14 sequências brasileiras disponíveis nos bancos de dados são muito próximas geneticamente, mas, comparativamente, a amostra sequenciada em 2020 é mais distante das demais. Este dado sugere uma possível evolução molecular do vírus ao longo dos anos, reforçando os achados de Vieira e colaboradores (2022). Especificamente quanto à amostra LPVA4, conforme esperado, esta foi agrupada próximo às sequências KR063107.1 e KR063108.1, isoladas na mesma região (Minas Gerais, Brasil) e ano (2015) (Vannucci *et al.*, 2015).

	MZ456812.1	MF615501.1	KR063107.1	KR063108.1	KR063109.1	MF615507.1	MF615510.1	MF615508.1	MF615509.1	MF615503.1	MF615504.1	MF615506.1	MF615505.1	MF615502.1
MZ456812.1														
MF615501.1	0.0203347357													
KR063107.1	0.0204238042	0.0025352462												
KR063108.1	0.0204238042	0.0025352462	0.0000000000											
KR063109.1	0.0202399507	0.0028869706	0.0031126243	0.0031126243										
MF615507.1	0.0217399573	0.0041316302	0.0043965010	0.0043965010	0.0038662143									
MF615510.1	0.0220223862	0.0044078137	0.0046735072	0.0046735072	0.0041415255	0.0002626885								
MF615508.1	0.0219276183	0.0043134367	0.0046607687	0.0046607687	0.0041302445	0.0005255653	0.0007886953							
MF615509.1	0.0218340672	0.0042260941	0.0044914010	0.0044914010	0.0039602249	0.0004379188	0.0007009787	0.0000874490						
MF615503.1	0.0213249696	0.0063761067	0.0068213317	0.0068213317	0.0062854000	0.0077295434	0.0079995843	0.0078981091	0.0078117633					
MF615504.1	0.0211388325	0.0063198539	0.0067683114	0.0067683114	0.0062295865	0.0076746091	0.0079458259	0.0078506985	0.0077605593	0.0000000000				
MF615506.1	0.0198822500	0.0053627329	0.0054489491	0.0054489491	0.0050950589	0.0065339884	0.0067288736	0.0066288660	0.0065434026	0.0048580626	0.0047948536			
MF615505.1	0.0223094986	0.0060896327	0.0065345232	0.0065345232	0.0060005014	0.0061863107	0.0064570044	0.0063593564	0.0062720581	0.0097037675	0.0095735304	0.0084209084		
MF615502.1	0.0238764494	0.0065371956	0.0069884787	0.0069884787	0.0062657643	0.0062693529	0.0065569087	0.0064659509	0.0063753766	0.0100505859	0.0100051075	0.0088127741	0.0076455721	

Figura 36 – Distância genética entre as 14 sequências brasileiras disponíveis no NCBI *Virus*.

Ressalta-se que a ausência de distância genética entre as amostras KR063107.1 e KR063108.1 é explicada pelo fato de que ambas são referentes ao mesmo animal, entretanto a primeira foi isolada de fluido vesicular, enquanto a segunda foi isolada do soro (informações indisponíveis no NCBI *Virus*, mas extraídas da publicação de Vannucci *et al.*, 2015). O mesmo argumento é válido para as sequências MF615503.1 e MF615504.1, obtidas a partir de amostra de pulmão e de vesícula na banda coronária, respectivamente (dados extraídos do NCBI *Virus*).

A figura 37, que relaciona o número de amostras sequenciadas com diferentes anos, demonstra que após os surtos de 2015 houve queda acentuada da quantidade de genomas de SVA sequenciados. Tendo em vista o perfil endêmico que a doença assumiu no Brasil e em

outros países (Brasil, 2020; Houston *et al.*, 2020), e o fato de que sua distribuição vem se ampliando para outras regiões, a exemplo do Chile, mais recentemente (Bennett *et al.*, 2022), a falta de sequenciamento não pode ser justificada pela menor ocorrência da doença. Assim, é importante alertar que a pesquisa contínua é necessária para que informações relevantes para maior compreensão e vigilância do vírus, independentemente do país, estejam constantemente disponíveis.

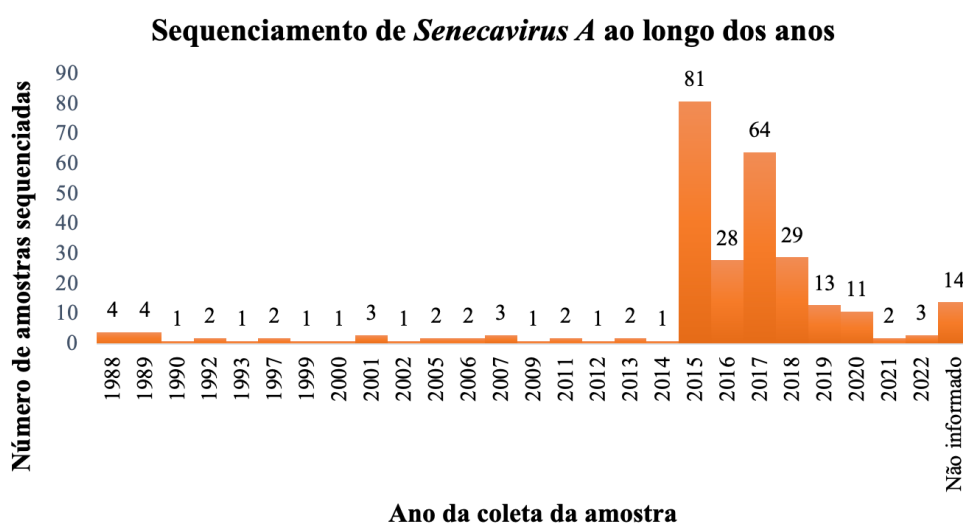


Figura 37 - Sequências completas de *Senecavirus A* disponibilizadas no NCBI *Virus* ao longo dos anos.

A atribuição de cor conforme o país de origem da amostra (figura 35), além de facilitar a visualização do agrupamento geográfico, permite observar a quantidade superior de sequências depositadas pelos Estados Unidos (em azul escuro) e China (em vermelho). O número de sequências depositadas por cada país está apresentado na figura 38, considerando apenas aquelas com tamanho próximo ao genoma completo (mais de 7000 pares de base), extraídas do NCBI *Virus* em 30/10/2023. Como os dados de prevalência de *Senecavirus A* são escassos no Brasil e no mundo, não é possível estabelecer uma relação entre o número de casos da doença em cada país e a geração de dados de sequenciamento, mesmo porque a incompatível disponibilidade de recursos financeiros para pesquisa em diferentes países não pode ser desconsiderada. Outro aspecto que deve ser ponderado é o poder que a suinocultura assume em diferentes economias. Neste contexto, China e Estados Unidos ocupam a primeira e terceira posições, respectivamente, no ranking dos principais produtores de suínos do mundo (ABPA, 2023), o que justifica a preocupação com patógenos de importância econômica, como é o caso do SVA. Ao seguir este raciocínio, fica evidente que o Brasil, quarto maior produtor e

exportador de carne suína (ABPA, 2023), deve buscar alternativas para ampliar o acesso às ferramentas de sequenciamento, seja mediante esforços políticos junto a instituições como o Centro Pan-Americano de Febre Aftosa e Saúde Pública Veterinária (PANAFTOSA), ao qual interessa o controle de SVA e de outros diagnósticos diferenciais de febre aftosa; ou por meio de parcerias com empresas privadas, que podem se beneficiar das informações fornecidas para desenvolvimento de produtos como vacinas, kits de diagnóstico e fármacos. Desta forma, o investimento no sequenciamento de amostras de SVA beneficia os diferentes elos do agronegócio brasileiro, sendo fundamental para avanços quanto à vigilância epidemiológica, prevenção e controle das infecções por *Senecavirus A*.

**Sequenciamento de amostras de *Senecavirus A* por país
(atribuição de cor conforme o número de sequências depositadas)**

1 6 14 117 124

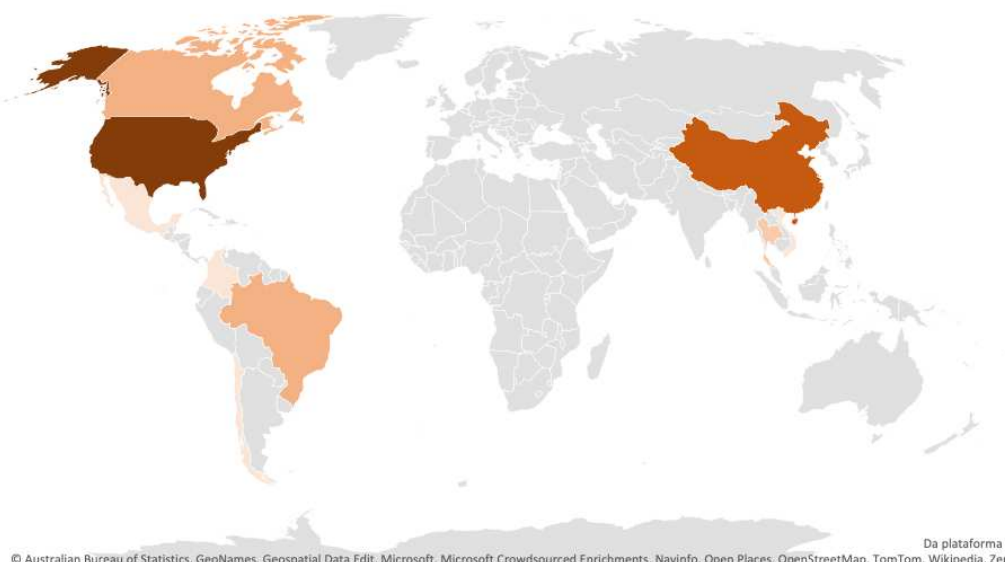


Figura 38 – Número de sequências completas de SVA disponibilizadas no NCBI *Virus* por diferentes países. Chile, Colômbia, México e Vietnã depositaram um genoma cada. A Tailândia depositou 6 genomas, seguidos de Brasil e Canadá, que contribuiriam para o banco de dados com 14 sequências. China e Estados Unidos são os países que mais sequenciaram, tendo depositado 117 e 124 sequências, respectivamente.

5.5 Predição de epítomos e avaliação da conservação

Como a montagem do genoma da amostra LPVA4 apresentou muitos *gaps*, e a utilização de alternativas para melhorar o resultado do sequenciamento e montagem foi postergada, optou-se por escolher outra sequência que estivesse disponível nos bancos de dados para a etapa de predição de epítomos. Dentre as opções, foi dada preferência para sequências

brasileiras isoladas em data próxima da sequência LPVA4, o que culminou na escolha da sequência SVA/BRA/MG1/2015 (acesso *Genbank*: KR063107), oriunda do mesmo estado (Minas Gerais) e isolada no mesmo ano (2015). A proximidade genética entre as duas sequências pode ser confirmada pela topologia da árvore construída, com a LPVA4 e a MG1 ocupando o mesmo clado (figura 35).

Após a busca por epítomos nas ferramentas disponíveis no IEDB e no *ABCPred*, filtragem e sobreposição dos determinantes antigênicos preditos para MHC I, MHC II e linfócitos B, foram selecionados 45 epítomos para construção da proteína M-E1, pretendida para uso em formulação vacinal (sequências não divulgadas). O alinhamento de 316 sequências de proteínas com tamanho próximo ao esperado para SVA (2181 aminoácidos) revelou boa conservação dos epítomos selecionados (dados não apresentados). Informações sobre as sequências utilizadas para a análise de conservação podem ser acessadas no Apêndice II.

A busca por epítomos de linfócitos B para fins diagnósticos, pela ferramenta *ABC Pred*, retornou inicialmente 70 epítomos com *score* maior ou igual a 0.80, referentes às proteínas *Leader*, 2B, 2C, 3A, 3C, VP1, VP2, VP3 e VP4 (dados não apresentados). A fim de reduzir o conjunto de epítomos e selecionar os mais antigênicos, foi alterado o parâmetro da busca para epítomos com *score* maior ou igual a 0.90, restringindo os resultados para 16 epítomos (dados não apresentados). Por fim, foi feita nova filtragem, para priorizar os epítomos com *score* maior ou igual a 0.90 referentes às proteínas do capsídeo (VP1, VP2, VP3 e VP4), devido à maior exposição destas às células de defesa *in vivo*. Assim, foram selecionados 10 epítomos de interesse, cujas informações estão disponíveis na tabela 3 (a sequência dos peptídeos foi omitida devido ao possível requerimento de patente). Dos 10 epítomos preditos, D2, D3, D4, D6, D7 e D9 coincidem parcialmente com determinantes antigênicos observados por outros grupos de pesquisa (Ru *et al.*, 2023; Zhang *et al.*, 2023).

Tabela 3 - Dados dos epítomos de linfócito B selecionados.

Epítopo	Proteína	Posição	Score
D1	VP1	80	0.96
D2	VP2	34	0.94
D3	VP2	220	0.93
D4	VP4	2	0.92
D5	VP2	239	0.92
D6	VP2	211	0.92
D7	VP2	158	0.92
D8	VP2	260	0.91
D9	VP1	162	0.91
D10	VP3	148	0.90

O alinhamento de proteínas das 14 sequências brasileiras disponíveis no NCBI *Virus* (códigos de acesso disponíveis no Apêndice III) revelou que, dos 10 epítomos selecionados, 7 estão conservados em todas as linhagens do Brasil, enquanto 3 epítomos (D1, D6 e D9) possuem alteração pontual de um aminoácido (indicada pelas setas) em todas as linhagens brasileiras (figura 39). Inicialmente, tem-se a perspectiva de sintetizar estes 7 peptídeos conservados em sintetizador automático para avaliar o potencial em testes diagnósticos diretos ou indiretos, individualmente e em combinação, semelhante ao descrito no trabalho de Santos e colaboradores (2022). De toda forma, os peptídeos D1 a D10 foram também utilizados para construção de proteína multi-epítopo (M-E2), para ampliar as possibilidades de aplicação destas moléculas.

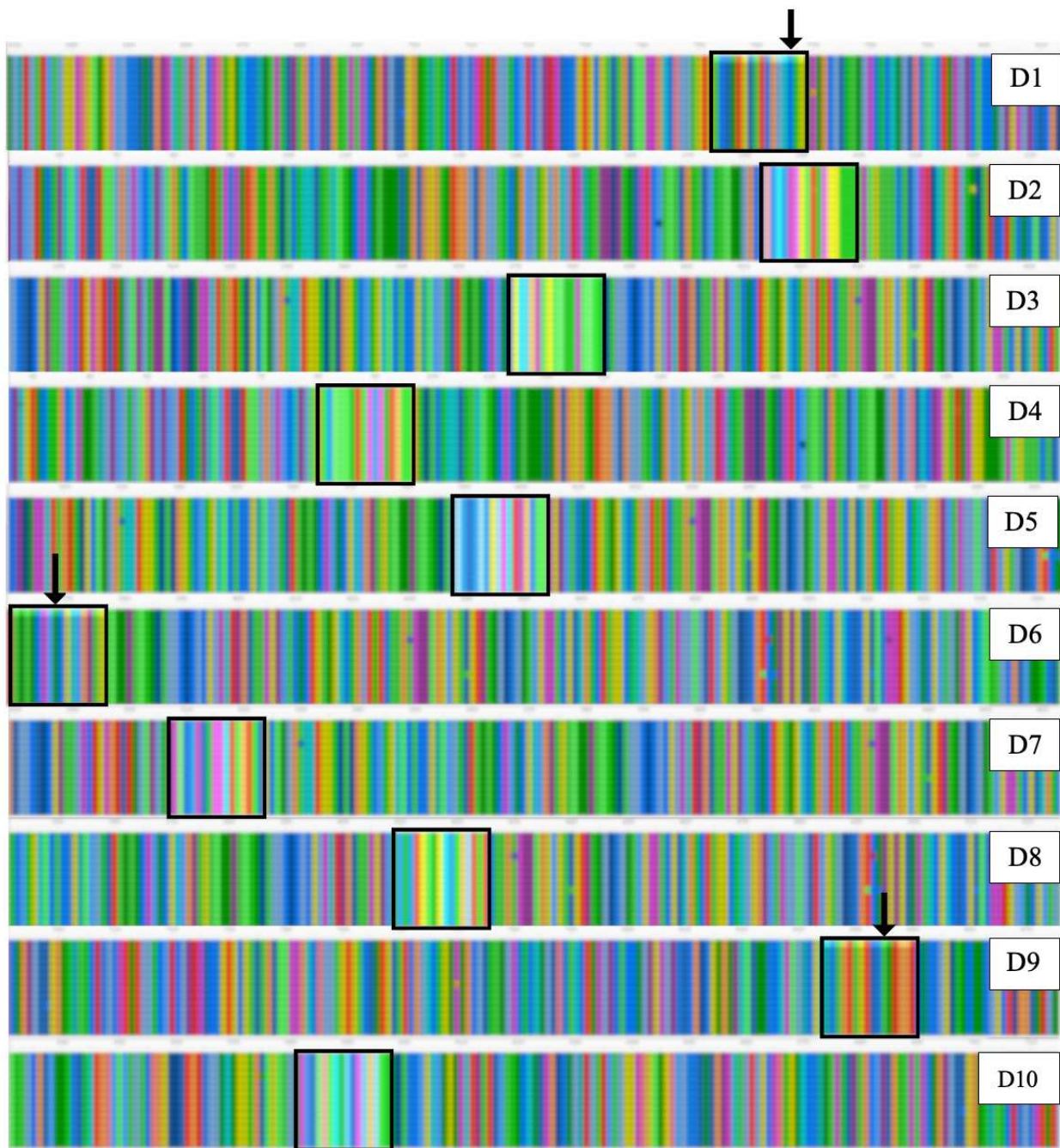


Figura 39 - Alinhamento das 14 sequências de proteínas de SVA oriundas do Brasil e disponíveis no NCBI *Virus*, com indicação da região de cada um dos dez peptídeos preditos (moldura preta), conforme identificação à direita. Cada cor na coluna representa um aminoácido diferente, e cada linha é uma sequência. A manutenção do mesmo padrão de cor em todas as linhas demonstra a conservação dos epítomos nas amostras brasileiras. As setas pretas apontam o sítio de alteração de aminoácido nos peptídeos D1, D6 e D9. Ressalta-se que a alteração apontada ocorre em todas as sequências brasileiras. Foi utilizado recurso de distorção da imagem a fim de manter a confidencialidade da sequência de aminoácidos de cada peptídeo.

5.6 Construção de proteínas multi-epítopo

Os 45 epítomos de linfócitos T CD4+/B e T CD8+/B foram conectados pelos *linkers* GPGPG e AAY, respectivamente. Os ligantes desempenham funções na expressão, dobramento, estabilização e atividade biológica da proteína (Vilela Rodrigues *et al.*, 2022). Ao todo, a proteína multi-epítopo M-E1, construída para ser testada como opção vacinal, possui 658 aminoácidos. Conforme destacado na metodologia, além da sequência pura, a proteína foi testada associada aos adjuvantes cadeia B da enterotoxina termolábil de *E. coli* (Sun *et al.*, 2013), GPX1 (Du *et al.*, 2017) e GP96 (Chen *et al.*, 2013), conectados no início da sequência pelo ligante EAAAK.

Do ponto de vista das estratégias *in silico* empregadas, a construção da proteína M-E1 se destaca para uso como antígeno vacinal em comparação com a proteína obtida por Zhang *et al.* (2023), uma vez que a M-E1 foi construída mediante associação de epítomos de linfócitos T CD4+, T CD8+ e B preditos especificamente a partir de genoma brasileiro de *Senecavirus A*; enquanto a proteína obtida no trabalho mencionado utilizou epítomos universais de linfócitos T, portanto, menos específicos. Por outro lado, cabe ressaltar que a vacina produzida por Zhang e colaboradores (2023) já obteve resultados promissores em experimentos *in vitro* e *in vivo*, que ainda são necessários para a proteína M-E1.

Para construção da proteína multi-epítopo M-E2, voltada para diagnóstico, foram testados três *linkers* descritos na literatura: GGSSGG (Zhang *et al.*, 2023), GSGSG (Shams *et al.*, 2021) e KK (Shawan *et al.*, 2023). Os melhores resultados, abordados no tópico 5.7, foram associados ao uso do ligante KK, de modo que a construção da M-E2 com este *linker* resultou em uma proteína com 178 aminoácidos.

5.7 Análises de antigenicidade, alergenicidade e toxicidade

Os resultados das análises de antigenicidade, alergenicidade e toxicidade da proteína multi-epítopo M-E1 estão apresentados na tabela 4.

Tabela 4 - Resultados das análises da proteína multi-epítopo M-E1, com e sem adjuvantes.

Parâmetro	Sem adjuvante	Cadeia B toxina <i>E. coli</i>	GPX1	GP96
Antigenicidade	0.5076	0.5059	0.5209	0.4082
Alergenicidade	Não	Não	Não	Não
Toxicidade	Não	1 peptídeo	3 peptídeos	11 peptídeos

Como a análise de toxicidade apontou a existência de peptídeos tóxicos nas construções vacinais associadas a adjuvantes, sem que houvesse incremento significativo na antigenicidade, optou-se por seguir as análises com a sequência da proteína sem associações. Neste contexto, na hipótese de futuramente serem realizados testes *in vivo*, será importante associar adjuvantes de uso bem relatado na suinocultura, a fim de intensificar o estímulo à resposta imune.

Em relação à proteína M-E2, os resultados estão apresentados na tabela 5. Nenhuma das construções testadas apresentou alergenicidade ou toxicidade. Devido à maior antigenicidade obtida pela utilização do *linker* KK, este foi escolhido para construção da proteína e foi a opção mantida para as demais análises.

Tabela 5 - Resultados das análises da proteína multi-epítopo M-E2, com diferentes opções de *linkers*,

Parâmetro	GGSSGG	GSGSG	KK
Antigenicidade	0.5100	0.5159	0.7032
Alergenicidade	Não	Não	Não
Toxicidade	Não	Não	Não

5.8 Avaliação de parâmetros físico-químicos e homologia com o hospedeiro

As análises dos parâmetros físico-químicos das proteínas M-E1 e M-E2 revelaram os resultados apresentados na tabela 6.

A proteína M-E1 é formada por 658 aminoácidos, possui massa molecular de 70.606,83 (70,61KD) e ponto isoelétrico 9,14, sugerindo o caráter básico da proteína. O coeficiente de extinção calculado ($127900 \text{ M}^{-1} \text{ cm}^{-1}$) deverá ser considerado como referência na hipótese de purificação da proteína por espectrofotômetro. O índice alifático de 72.87 indica uma boa estabilidade da proteína diante de mudanças de temperatura. A meia-vida calculada é de 30 horas em reticulócitos de mamíferos (*in vitro*), >20 horas em leveduras (*in vivo*) e >10 horas em *E. coli* (*in vivo*), o que já foi considerado como um resultado promissor para candidatos vacinais em outros estudos (Vilela Rodrigues *et al.*, 2022). O índice de instabilidade (II), calculado em 26.63, indica a estabilidade da proteína. O GRAVY calculado é de 0,018, valor positivo que sugere o caráter hidrofóbico da proteína. Por outro lado, o resultado da ferramenta *Protein-Sol* (0,289) sugere solubilidade da proteína, ainda que inferior à média das proteínas de *E. coli* (0,45) (Hebditch *et al.*, 2017; Vilela Rodrigues *et al.*, 2022). Os resultados obtidos indicam um bom potencial para o uso da proteína como antígeno vacinal, sendo suficientes para justificar a realização futura de testes *in vitro* e *in vivo*.

Tabela 6 - Parâmetros físico-químicos das proteínas multi-epítipo M-E1 e M-E2.

Parâmetro	M-E1	M-E2
Nº aminoácidos	658	178
Peso molecular	70606.83	19975.58
pI	9.14	9.85
Coefficiente extinção	127900	23045/22920
Índice alifático	72.87	44.89
Meia-vida estimada	30h/>20/>10	1.4h/3min/>10h
Índice de instabilidade	26.63 (estável)	30.35 (estável)
GRAVY	0.018	-1.046
Solubilidade	0.289	0.698

Quanto à proteína M-E2, esta é formada por 178 aminoácidos, possui massa molecular de 19.975,58 (19,98KD) e caráter básico, evidenciado pelo ponto isoelétrico de 9,85. Foram calculados dois coeficientes de extinção (23045 e 22920 $M^{-1} cm^{-1}$), que devem ser considerados para eventual purificação da proteína. O primeiro valor calculado assume que os pares de cisteína formam cistinas, enquanto o segundo considera que todos os resíduos de cisteína sofrem redução. O índice alifático de 44.89 e o índice de instabilidade de 30.35 sugerem a estabilidade da proteína. O GRAVY de -0.146 indica que a proteína é hidrofílica, com solubilidade de 0,689 calculada pelo *Protein-Sol*. O resultado da meia-vida para a proteína M-E2 foi inferior ao obtido para a M-E1, com estimativa de 1.4 horas de meia-vida em reticulócitos de mamíferos (*in vitro*), 3 minutos em leveduras (*in vivo*) e > 10 horas em *E. coli* (*in vivo*). Apesar de este resultado estar abaixo do esperado para mamíferos e leveduras, a meia-vida superior a 10 horas em *E. coli*, provável vetor para a expressão da proteína, permite supor que a M-E2 seja adequada para fins diagnósticos. Outras análises em perspectiva, como a de simulação dinâmica, poderão esclarecer mais detalhes a respeito da funcionalidade da proteína, assim como futuros testes *in vitro*. Desta forma, o resultado de meia-vida estimada, avaliado em conjunto com os demais resultados e pensado no contexto flexível de predições *in silico*, não invalida a perspectiva de uso da M-E2 em testes diagnósticos.

5.9 Análises de homologia

Em relação à análise de homologia com o hospedeiro (*Sus scrofa domesticus*), nenhuma proteína similar à M-E1 (figura 40-A) ou M-E2 (figura 40-B) foi identificada. Assim, não há

indício de que as proteínas multi-epítipo construídas possam desencadear reações de autoimunidade em suínos domésticos ou reações inespecíficas em métodos diagnósticos com amostras de suínos.

A

Your search is limited to records that include: *Sus scrofa domesticus* (taxid:9825)

Job Title	Multi-epitope protein M-E1
RID	UUDXY1U9013 <small>Search expires on 01-22 22:26 pm</small> Download All ▼
Program	Citation ▼
Database	nr See details ▼
Query ID	lcl Query_2036654
Description	unnamed protein product
Molecule type	amino acid
Query Length	658
Other reports	?

Filter Results

Percent Identity	E value	Query Coverage
<input type="text"/> to <input type="text"/>	<input type="text"/> to <input type="text"/>	<input type="text"/> to <input type="text"/>
<input type="button" value="Filter"/> <input type="button" value="Reset"/>		

⚠ No significant similarity found. For reasons [why](#), [click here](#)

B

Your search is limited to records that include: *Sus scrofa domesticus* (taxid:9825)

Job Title	Multi-epitope protein M-E2
RID	UJUEYCVX2013 <small>Search expires on 01-22 22:44 pm</small> Download All ▼
Program	Citation ▼
Database	nr See details ▼
Query ID	lcl Query_2852441
Description	unnamed protein product
Molecule type	amino acid
Query Length	178
Other reports	?

Filter Results

Percent Identity	E value	Query Coverage
<input type="text"/> to <input type="text"/>	<input type="text"/> to <input type="text"/>	<input type="text"/> to <input type="text"/>
<input type="button" value="Filter"/> <input type="button" value="Reset"/>		

⚠ No significant similarity found. For reasons [why](#), [click here](#)

Figura 40 - Resultado do BLASTp realizado para verificar a existência de homologia entre o proteoma de suínos domésticos e as proteínas M-E1 (A) e M-E2 (B). Não foram identificadas proteínas similares.

Quanto à homologia entre as proteínas construídas e o proteoma dos vírus da febre aftosa, estomatite vesicular, doença vesicular dos suínos e exantema vesicular dos suínos, não foi verificada similaridade significativa entre estes e a proteína M-E1 (figura 41-A a 41-D). Entretanto, foi verificada homologia parcial entre a proteína M-E2 e 94 sequências do proteoma do vírus da doença vesicular dos suínos (figura 42-C), com identidades de até 47,83%. Apesar da correspondência encontrada, o que inicialmente preocupa quanto à especificidade diagnóstica diante do uso da proteína, como se trata de uma doença ainda considerada exótica no Brasil o maior prejuízo seria quanto à não detecção deste vírus e alerta para sua entrada no país. Todavia, esta preocupação não merece prosperar, uma vez que a doença vesicular dos suínos é de baixa incidência mundial, registrada apenas em países europeus e asiáticos; mesmo que houvesse demora para confirmar a entrada do agente, as medidas de controle são baseadas em princípios de biossegurança, da mesma forma que o controle de SVA; e não haveria

implicação quanto ao diagnóstico diferencial de maior importância, febre aftosa, que de fato exigiria as restrições comerciais e prejuízos econômicos discutidos previamente (Brasil, 2020). De toda forma, a avaliação *in vitro* pretendida para proteína M-E2, bem como para os epítomos de linfócito B preditos, isoladamente e em pares, é a ferramenta que irá responder efetivamente quanto à real possibilidade de interferência diagnóstica. Só a partir destes resultados será possível avaliar se os epítomos preditos e a proteína M-E2 poderão servir ao propósito de promover um diagnóstico sensível e específico de *Senecavirus A*.

A Your search is limited to records that include: Foot-and-mouth disease virus (taxid:12110)

Job Title: Multi-epitope protein M-E1
 RID: [UWFY51BH016](#) Search expires on 01-23 17:13 pm [Download All](#) ▾
 Program: [Citation](#) ▾
 Database: nr [See details](#) ▾
 Query ID: Icl|Query_11226539
 Description: unnamed protein product
 Molecule type: amino acid
 Query Length: 658
 Other reports: [?](#)

Filter Results
 Percent Identity: to
 E value: to
 Query Coverage: to
[Filter](#) [Reset](#)

⚠ No significant similarity found. For reasons why [click here](#)

B Your search is limited to records that include: Vesicular stomatitis virus (taxid:11276)

Job Title: Multi-epitope protein M-E1
 RID: [UWGD33DN013](#) Search expires on 01-23 17:21 pm [Download All](#) ▾
 Program: [Citation](#) ▾
 Database: nr [See details](#) ▾
 Query ID: Icl|Query_709916
 Description: unnamed protein product
 Molecule type: amino acid
 Query Length: 658
 Other reports: [?](#)

Filter Results
 Percent Identity: to
 E value: to
 Query Coverage: to
[Filter](#) [Reset](#)

⚠ No significant similarity found. For reasons why [click here](#)

C Your search is limited to records that include: Swine vesicular disease virus (taxid:12075)

Job Title: Multi-epitope protein M-E1
 RID: [UWG7CD04016](#) Search expires on 01-23 17:18 pm [Download All](#) ▾
 Program: [Citation](#) ▾
 Database: nr [See details](#) ▾
 Query ID: Icl|Query_330410
 Description: unnamed protein product
 Molecule type: amino acid
 Query Length: 658
 Other reports: [?](#)

Filter Results
 Percent Identity: to
 E value: to
 Query Coverage: to
[Filter](#) [Reset](#)

⚠ No significant similarity found. For reasons why [click here](#)

D Your search is limited to records that include: Vesicular exanthema of swine virus (taxid:35612)

Job Title: Multi-epitope protein M-E1
 RID: [UWG44S4F013](#) Search expires on 01-23 17:16 pm [Download All](#) ▾
 Program: [Citation](#) ▾
 Database: nr [See details](#) ▾
 Query ID: Icl|Query_419003
 Description: unnamed protein product
 Molecule type: amino acid
 Query Length: 658
 Other reports: [?](#)

Filter Results
 Percent Identity: to
 E value: to
 Query Coverage: to
[Filter](#) [Reset](#)

⚠ No significant similarity found. For reasons why [click here](#)

Figura 41 – Análise de homologia entre a proteína M-E1 e o vírus da febre aftosa (A), estomatite vesicular (B), doença vesicular dos suínos (C) e exantema vesicular dos suínos (D). Não foi verificada similaridade significativa.

A **i** Your search is limited to records that include: Foot-and-mouth disease virus (taxid:12110)

Job Title **Multi-epitope protein M-E2**

RID [UWGG34ZR016](#) Search expires on 01-23 17:22 pm [Download All](#) ▼

Program [Citation](#) ▼

Database nr [See details](#) ▼

Query ID IcllQuery_787070

Description unnamed protein product

Molecule type amino acid

Query Length 178

Other reports [?](#)

Filter Results

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

A No significant similarity found. For reasons why, [click here](#)

B **i** Your search is limited to records that include: Vesicular stomatitis virus (taxid:11276)

Job Title **Multi-epitope protein M-E2**

RID [UWGXBCV2013](#) Search expires on 01-23 17:30 pm [Download All](#) ▼

Program [Citation](#) ▼

Database nr [See details](#) ▼

Query ID IcllQuery_1133966

Description unnamed protein product

Molecule type amino acid

Query Length 178

Other reports [?](#)

Filter Results

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

A No significant similarity found. For reasons why, [click here](#)

C **i** Your search is limited to records that include: Swine vesicular disease virus (taxid:12075)

Job Title **Multi-epitope protein M-E2**

RID [UWGMFHD016](#) Search expires on 01-23 17:25 pm [Download All](#) ▼

Program BLASTP [Citation](#) ▼

Database nr [See details](#) ▼

Query ID IcllQuery_923570

Description unnamed protein product

Molecule type amino acid

Query Length 178

Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

Filter Results

Organism *only top 20 will appear* exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100 [?](#)

select all 94 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> polyprotein [Swine vesicular disease virus]	Swine vesicular di...	29.3	29.3	12%	0.029	47.83%	2185	ALN12393.1
<input checked="" type="checkbox"/> Chain C_Polyprotein Capsid Protein [Swine vesicular disease virus]	Swine vesicular di...	30.0	30.0	25%	0.012	36.96%	238	1MQT_C
<input checked="" type="checkbox"/> Chain C_Coat protein VP3 [Swine vesicular disease virus (STRAIN UKG/27/72)]	Swine vesicular di...	30.0	30.0	25%	0.012	36.96%	238	1OOP_C

D **i** Your search is limited to records that include: Vesicular exanthema of swine virus (taxid:35612)

Job Title **Multi-epitope protein M-E2**

RID [UWG44S4F013](#) Search expires on 01-23 17:16 pm [Download All](#) ▼

Program [Citation](#) ▼

Database nr [See details](#) ▼

Query ID IcllQuery_419003

Description unnamed protein product

Molecule type amino acid

Query Length 658

Other reports [?](#)

Filter Results

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

A No significant similarity found. For reasons why, [click here](#)

Figura 42 - Análise de homologia entre a proteína M-E2 e o vírus da febre aftosa (A), estomatite vesicular (B), doença vesicular dos suínos (C) e exantema vesicular dos suínos (D). Foi identificada correspondência com 94 sequências do proteoma do vírus da doença vesicular dos suínos.

5.10 Análises estruturais

A estrutura secundária da proteína M-E1 está representada na figura 43. A estrutura terciária inicial, anterior ao refinamento, está ilustrada na figura 44, e o *Ramachandran plot* correspondente na figura 45.

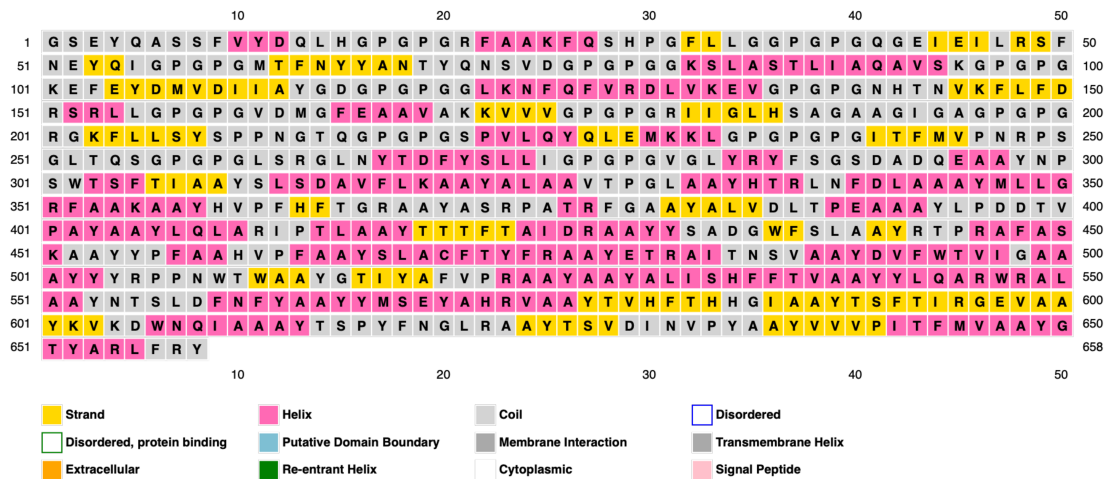
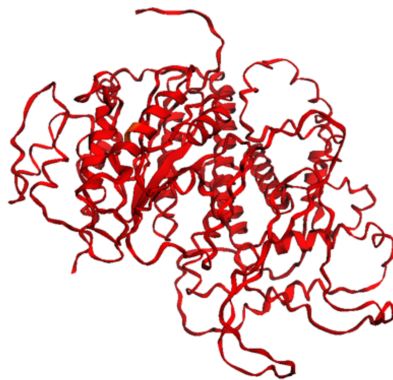


Figura 43 - Estrutura secundária da proteína M-E1, dada pelo PSIPRED 4.0, indicando as regiões de folha β (amarelo), α -hélice (rosa) e bobinas (cinza).



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

Figura 44 - Estrutura terciária inicialmente predita pelo *AlphaFold2* para a proteína M-E1.

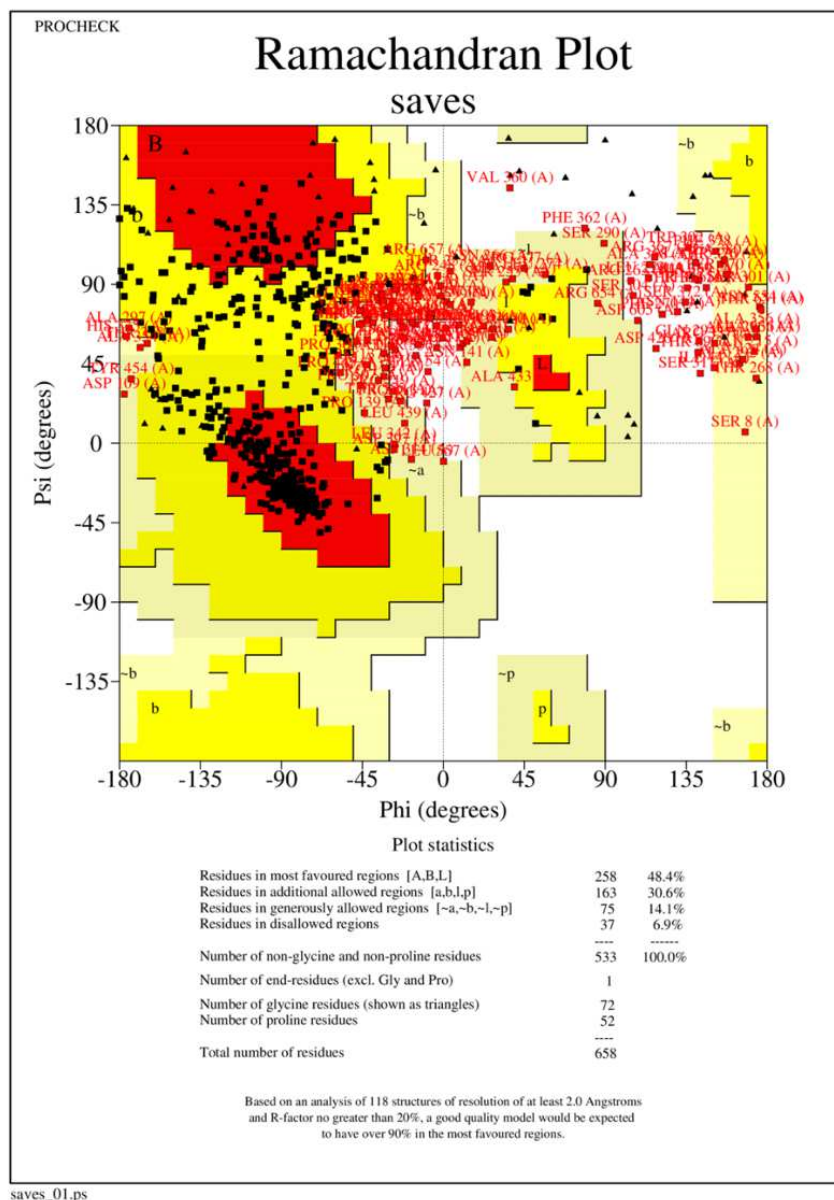
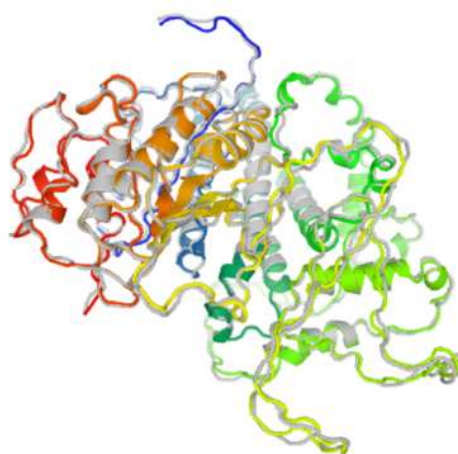


Figura 45 - *Ramachandran plot* da estrutura inicialmente predita para a proteína M-E1, gerado pelo programa PROCHECK, na qual apenas 48,4% dos resíduos estavam localizados nas regiões mais favoráveis.

Os resultados iniciais obtidos para a estrutura terciária da proteína M-E1 foram insatisfatórios. A estrutura predita pelo *Alphafold2* (figura 44) revela que todas as regiões da proteína foram modeladas com confiabilidade muito baixa (pLDDT < 50), o que sugere desordem estrutural (Tunyasuvunakool *et al.*, 2021). Além disso, o *Ramachandran plot* (figura 45) aponta que apenas 48,4% dos resíduos estavam em regiões favoráveis, com ângulos de torção viáveis; enquanto 30,6% estão em regiões permitidas, 14,1% em regiões generosamente permitidas (ou seja, permitidas quando analisadas com menor rigor) e 6,9% estão em regiões proibidas, nas quais os ângulos de torção assumem valores incompatíveis estereoquimicamente,

havendo risco de colisão entre átomos e comprometimento da funcionalidade da proteína (Birbeck College, [s.d.]; Laskowski *et al.*, 1993).

O refinamento da proteína pelo *GalaxyRefine* resultou na predição de cinco modelos tridimensionais, dos quais o modelo 1 apresentou melhor qualidade, com 95,3% dos resíduos em regiões favoráveis (figura 46). O *Ramachandran plot* referente à estrutura refinada (figura 47) evidenciou a presença de 92,7% dos resíduos em regiões favoráveis, 6,4% em regiões permitidas, 0,6% em regiões generosamente permitidas e 0,4% em regiões proibidas. Assim, a estrutura refinada atende ao padrão necessário para uma estrutura tridimensional de qualidade, na qual ao menos 90% dos resíduos devem estar nas regiões mais favoráveis (Birbeck College, [s.d.]; Laskowski *et al.*, 1993).



Structure Information

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	4.059	65.9	11.7	52.6
MODEL 1	0.8302	0.751	1.691	6.6	0.0	95.3
MODEL 2	0.8560	0.651	1.682	5.8	0.2	94.7
MODEL 3	0.8549	0.659	1.703	6.0	0.2	94.5
MODEL 4	0.8587	0.644	1.719	6.0	0.0	94.2
MODEL 5	0.8321	0.760	1.654	5.8	0.0	95.1

Figura 46 - Estrutura terciária da proteína M-E1, após refinamento pelo *GalaxyRefine*. Dos cinco modelos preditos, o modelo 1 teve o melhor resultado, com 95,3% dos resíduos em regiões favoráveis.

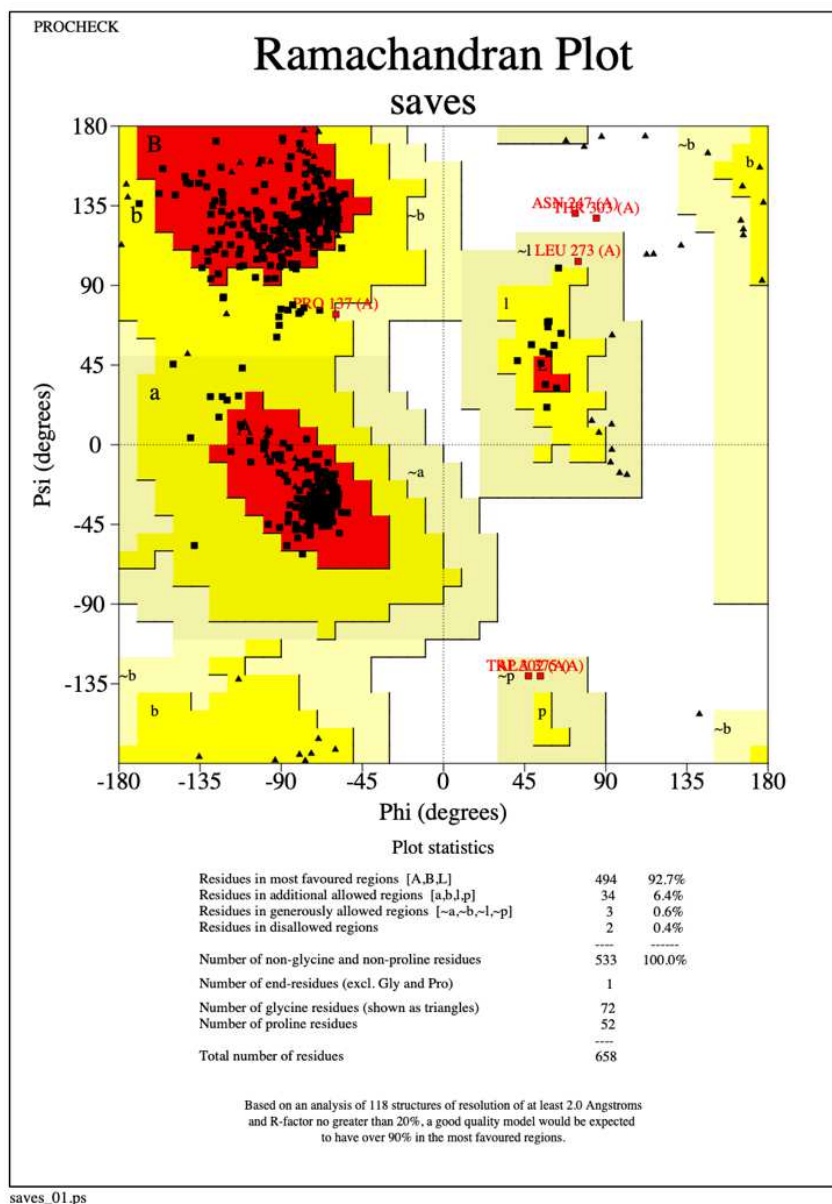


Figura 47 - *Ramachandran plot* da estrutura refinada da proteína M-E1, gerado pelo programa PROCHECK, na qual 92,7% dos resíduos estão localizados nas regiões mais favoráveis, sugerindo uma estrutura tridimensional de qualidade.

A estrutura secundária da proteína M-E2 apresenta regiões de alfa-hélice, folhas beta e bobinas, semelhante ao predito para a proteína M-E1 (figura 48). A predição tridimensional inicial (figura 49), dada pelo *AlphaFold2*, revelou regiões de confiabilidade muito baixa (pLDDT < 50) a razoável (pLDDT entre 70 e 80). O *Ramachandran plot* correspondente (figura 50) demonstrou que 29,6% dos resíduos estavam em regiões favoráveis, 47,4% em regiões permitidas, 17,8% em regiões generosamente permitidas e 5,3% em regiões proibidas.

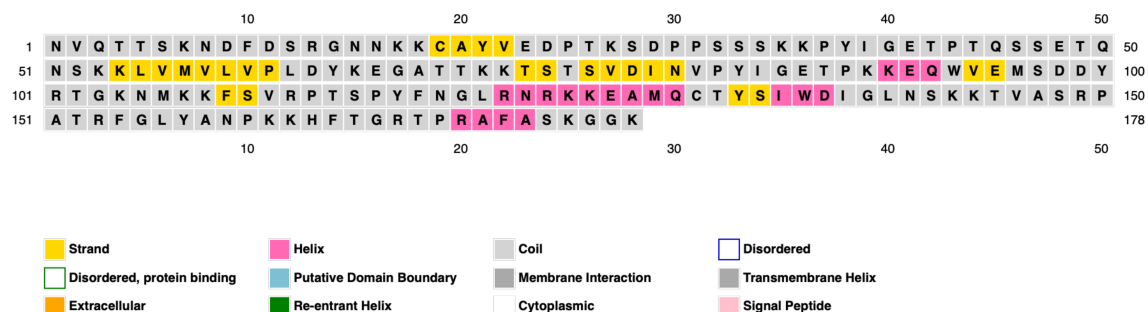
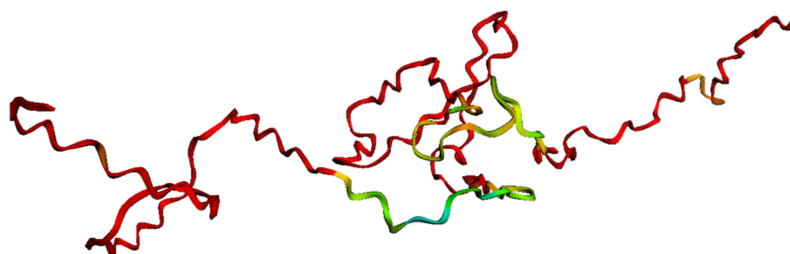


Figura 48 - Estrutura secundária da proteína M-E2, dada pelo PSIPRED 4.0, indicando as regiões de folha β (amarelo), α -hélice (rosa) e bobinas (cinza).



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

Figura 49 - Estrutura terciária inicialmente prevista pelo *AlphaFold2* para a proteína M-E2.

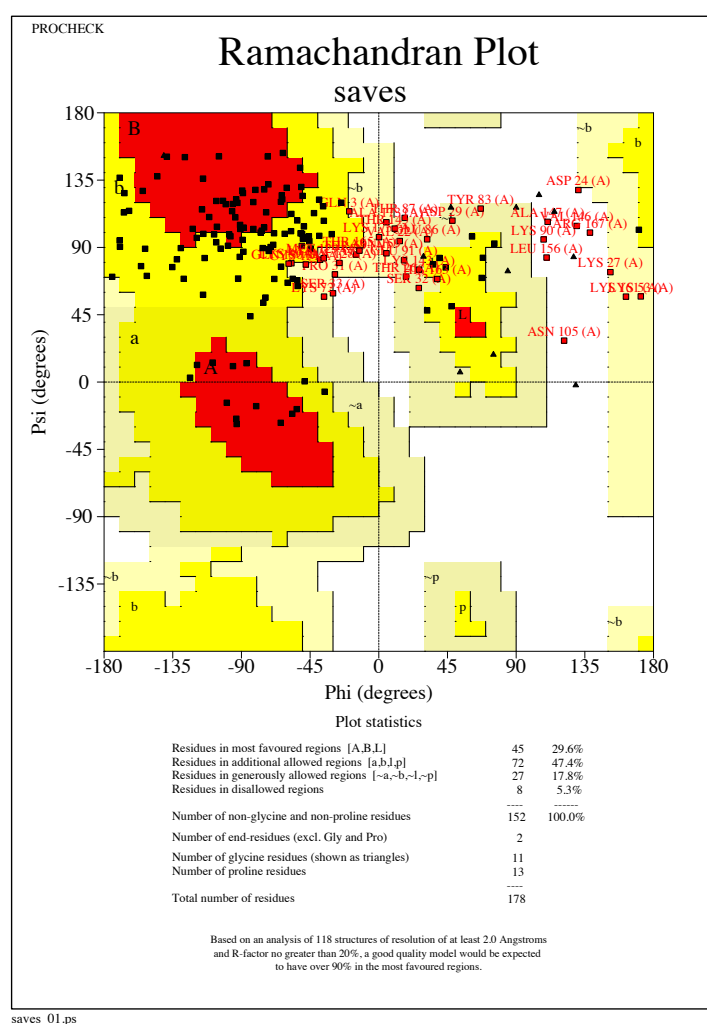
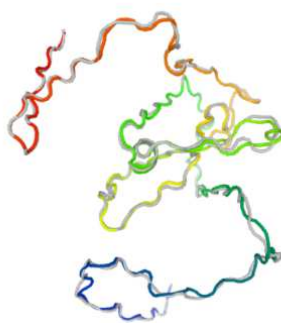


Figura 50 - *Ramachandran plot* da estrutura inicialmente predita para a proteína M-E2, gerado pelo programa PROCHECK, na qual apenas 29,6% dos resíduos estavam localizados nas regiões mais favoráveis.

O modelo 1 obtido pelo *GalaxyRefine* após o refinamento da proteína M-E2 apresentou 96% dos resíduos em regiões favoráveis (figura 51). Já o *Ramachandran plot* (figura 52), após o refinamento, indicou a presença de 89,5% de resíduos em áreas mais recomendadas, valor discretamente abaixo do limite de 90% para uma estrutura ser considerada de qualidade (Birbeck College, [s.d.]; Laskowski *et al.*, 1993). Além disso, foram identificados 7,9% dos resíduos em regiões permitidas, 0,7% em regiões generosamente permitidas e 2% em regiões proibidas. Desta forma, a estrutura refinada apresenta uma qualidade aceitável, mas que pode ser aprimorada por novos refinamentos.



Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.845	45.2	8.2	39.8
MODEL 1	0.8132	0.807	1.442	3.8	0.0	96.0
MODEL 2	0.7978	0.842	1.593	5.2	0.0	95.5
MODEL 3	0.7935	0.874	1.365	2.1	0.0	94.3
MODEL 4	0.7851	0.869	1.187	1.4	0.0	95.5
MODEL 5	0.7949	0.895	1.456	3.1	0.6	94.9

Figura 51 - Estrutura terciária da proteína M-E2, após refinamento pelo *GalaxyRefine*. Dos cinco modelos preditos, o modelo 1 teve o melhor resultado, com 96% dos resíduos em regiões favoráveis.

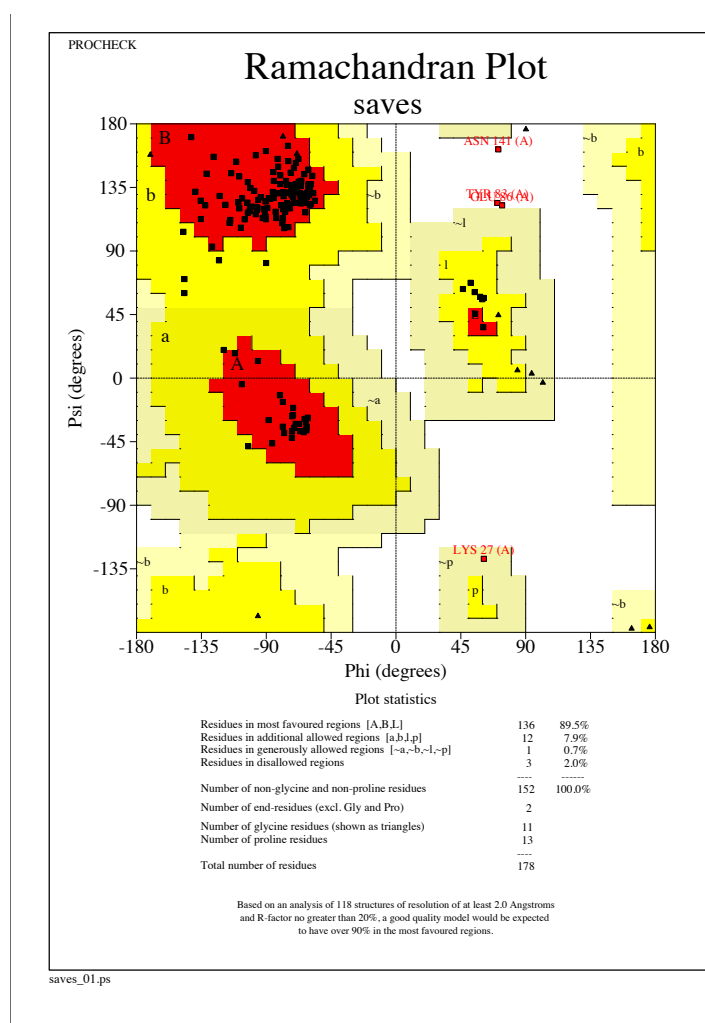


Figura 52 - *Ramachandran plot* da estrutura refinada da proteína M-E2, gerado pelo programa PROCHECK, na qual 89,5% dos resíduos estão localizados nas regiões mais favoráveis.

5.11 Predição de epítomos conformacionais de células B

A busca por epítomos conformacionais de linfócitos B no programa *ElliPro* resultou em 25 epítomos conformacionais identificados para a M-E1, sendo 10 deles com *score* acima de 0,80, ou seja, com confiabilidade apropriada da predição (dados não apresentados). Quanto à proteína M-E2, foram encontrados 9 epítomos conformacionais, 3 com *score* superior a 0,80. Os epítomos conformacionais são aqueles associados ao dobramento tridimensional da proteína, favorecendo a imunogenicidade para além da promovida pelos epítomos lineares (Lo *et al.*, 2021). Assim, a identificação dos epítomos conformacionais nas duas proteínas construídas reforça os resultados promissores encontrados nas demais análises realizadas, atestando o potencial de utilização das proteínas M-E1 e M-E2 para fins vacinais e de diagnóstico, respectivamente, a ser validado futuramente pelas análises *in vitro* e *in vivo* cabíveis.

6 CONCLUSÃO

O sequenciamento, alinhamento e análise filogenética realizados favoreceram a compreensão e ensejaram reflexões relevantes a respeito da epidemiologia molecular de *Senecavirus A*. As informações apresentadas indicam que esforços devem ser feitos por parte do Brasil e de outros países para manter os bancos de dados genômicos e proteômicos atualizados. As duas proteínas multi-epítopo construídas apresentaram resultados promissores nas análises *in silico*, demonstrando potencial para uso em formulação vacinal e métodos diagnósticos. Da mesma forma, foi possível predizer epítomos lineares de linfócitos B para fins diagnósticos, com alta pontuação, com indícios favoráveis de que possam ser úteis em ensaios sorológicos. Assim, testes *in vitro* e *in vivo* devem ser conduzidos para confirmar as predições obtidas por ferramentas de bioinformática.

7 PERSPECTIVAS

Diante dos resultados promissores encontrados durante o desenvolvimento deste trabalho, surgem perspectivas para a continuidade da pesquisa. Um novo sequenciamento da amostra LPVA4 pode ser realizado futuramente, a partir do uso de amplicons para melhorar a qualidade do sequenciamento, ou de plataformas de terceira geração. Uma nova montagem, seguida de anotação funcional de genes e proteínas, pode ser conduzida, com disponibilização da sequência nos bancos de dados genômicos. Análises mais complexas de bioinformática como as de *docking* molecular, simulação dinâmica, seleção positiva, resposta imune e clonagem *in silico* podem aumentar a confiabilidade quanto à imunogenicidade e

funcionalidade da proteína. Após a síntese dos peptídeos preditos, por sintetizador automático, bem como após a expressão recombinante das proteínas multi-epítipo, poderão ser efetuados testes *in vitro* e/ou *in vivo* necessários para confirmar as predições e conferir maior robustez aos resultados obtidos até então.

8 REFERÊNCIAS

ABBAS, Abul K.; LICHTMAN, Andrew H.; PILLAI, Shiv. **Cellular and Molecular Immunology**. 10. ed. Filadélfia: Elsevier, 2022.

ABPA – Associação Brasileira de Proteína Animal. Relatório anual. 2023. Disponível em: <<https://abpa-br.org/wp-content/uploads/2023/04/Relatorio-Anual-2023.pdf>>. Acesso em: 08/12/2023.

ADIYAMAN, R.; MCGUFFIN, L. Methods for the Refinement of Protein Structure 3D Models. **International Journal of Molecular Sciences**, [s. l.], v. 20, n. 9, p. 2301, 2019.

ALEXANDERSEN, S. *et al.* Picornaviruses. In: ZIMMERMAN, J. J. (editor) *et al.* Diseases of Swine. 11 ed. Hoboken, NJ : Wiley-Blackwell, 2019, p. 673-675

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, [s. l.], v. 215, n. 3, p. 403–410, 1990.

ANDREWS, Simon. FastQC A Quality Control tool for High Throughput Sequence Data. 2014. Disponível em: <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>>. Acesso em: 04/01/2024.

ARZT, J.; BERTRAM, M. R.; VU, L. T.; PAUSZEK, S. J.; HARTWIG, E. J.; SMOLIGA, G. R.; PALINSKI, R.; STENFELDT, C.; FISH, I. H.; HOANG, B. H.; PHUONG, N. T.; HUNG, V. V.; VU, P. P.; DUNG, N. K.; DONG, P. V.; TIEN, N. N.; DUNG, D. H. First Detection and Genome Sequence of Senecavirus A in Vietnam. **Microbiology Resource Announcements**, [s. l.], v. 8, n. 3, p. e01247-18, 2019.

BABRAHAM BIOINFORMATICS. FasqQC Help – Analysis Modules. [s.d.], a. Disponível em: <<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3%20Analysis%20Modules/>>. Acesso em: 10/01/2024.

BABRAHAM BIOINFORMATICS. FastQC Report – Good *Illumina* results. [s. d.], b. Disponível em: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/good_sequence_short_fastqc.html>. Acesso em: 10/01/2024.

BAHRAMI, A. A.; PAYANDEH, Z.; KHALILI, S.; ZAKERI, A.; BANDEHPOUR, M. Immunoinformatics: *In Silico* Approaches and Computational Design of a Multi-epitope, Immunogenic Protein. **International Reviews of Immunology**, [s. l.], v. 38, n. 6, p. 307–322, 2019.

BAI, M.; WANG, R.; SUN, S.; ZHANG, Y.; DONG, H.; GUO, H. Development and validation of a competitive ELISA based on virus-like particles of serotype Senecavirus A to detect serum antibodies. **AMB Express**, [s. l.], v. 11, n. 1, p. 7, 2021.

BALBINOT, Eduardo. Workflow científico de anotação genômica funcional e curadoria manual de genomas 2020. Dissertação (Mestrado) - Universidade de Caxias do Sul, Programa de Pós-Graduação em Biotecnologia, 2020.

BECKER, D. L.; ZANELLA, J. R. C.; CORBELLINI, L. G.; BORBA, M. R. Analysis of the Performance of the Animal Health Surveillance System in the Outbreak of Swine Vesicular Disease in the State of Santa Catarina - Brazil. **Acta Scientiae Veterinariae**, [s. l.], v. 48, 2020. Disponível em: <<https://seer.ufrgs.br/ActaScientiaeVeterinariae/article/view/103772>>. Acesso em: 24 dez. 2023.

BENNETT, B.; URZÚA-ENCINA, C.; PARDO-ROA, C.; ARIYAMA, N.; LECOCQ, C.; RIVERA, C.; BADÍA, C.; SUÁREZ, P.; AGREDO, M.; AGUAYO, C.; ÁVILA, C.; ARAYA, H.; PÉREZ, P.; BERRIOS, F.; AGÜERO, B.; MENDIETA, V.; PITUCO, E. M.; DE ALMEIDA, I. G.; MEDINA, R.; BRITO, B.; JOHOW, M.; RAMIREZ, V. N. First report and genetic characterization of Seneca Valley virus (SVV) in Chile. **Transboundary and Emerging Diseases**, [s. l.], v. 69, n. 6, 2022. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/tbed.14747>>. Acesso em: 7 dez. 2023.

BILIĆ, P.; KULEŠ, J.; GALAN, A.; GOMES DE PONTES, L.; GUILLEMIN, N.; HORVATIĆ, A.; FESTA SABES, A.; MRLJAK, V.; ECKERSALL, P. D. Proteomics in Veterinary Medicine and Animal Science: Neglected Scientific Opportunities with Immediate Impact. **PROTEOMICS**, [s. l.], v. 18, n. 14, p. 1800047, 2018.

BIRBECK COLLEGE. Ramachandran plot. [s.d.]. Disponível em: <https://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.html>. Acesso em: 12/01/2024.

BRACHT, A. J.; O'HEARN, E. S.; FABIAN, A. W.; BARRETTE, R. W.; SAYED, A. Real-Time Reverse Transcription PCR Assay for Detection of Senecavirus A in Swine Vesicular Diagnostic Specimens. **PLOS ONE**, [s. l.], v. 11, n. 1, p. e0146211, 2016.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Plano Estratégico do PNEFA - 2017 – 2026. Brasília, DF: v. 1.0,52 p., mar. 2017. Disponível em: <www.gov.br/agricultura/pt-br/assuntos/sanidade-animal-e-vegetal/saude-animal/programas-de-saude-animal/febre-aftosa/FEBREAFTOSAV6.pdf>. Acesso em: 08/12/2023.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Manual de investigação de doença vesicular. 1a ed. 60p. Brasília, DF: MAPA/AECS, 2020. Disponível em: <https://www.agrodefesa.go.gov.br/files/PEEV/Manual_12413089_manual_investigacao_doenca_vesicular.pdf>. Acesso em: 08/12/2022.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Manual de métodos oficiais para diagnóstico de doenças animais. 1a ed. Brasília, DF: MAPA/DAS/CGAL, 2022. Disponível em: <<https://www.gov.br/agricultura/pt-br/aceso-a-informacao/participacao-social/consultas-publicas/2022/arquivos-das-consultas-publicas/MANUALDIACONSULTAPUBLICA1.pdf>>. Acesso em: 11/03/2024.

BRAUN, R. L.; JUNQUEIRA, D. M.; VERLI, H. Filogenia molecular. In: VERLI, Hugo. *Bioinformática: da biologia à flexibilidade molecular*. 1. ed. São Paulo: SBBq, 2014.

BRUNO, L.; CORTESE, M.; RAPPUOLI, R.; MEROLA, M. Lessons from Reverse Vaccinology for viral vaccine design. **Current Opinion in Virology**, Viral pathogenesis • Preventive and therapeutic vaccines. [s. l.], v. 11, Viral pathogenesis • Preventive and therapeutic vaccines, p. 89–97, 2015.

BUCKLEY, A. C.; MICHAEL, D. D.; FAABERG, K. S.; GUO, B.; YOON, K.-J.; LAGER, K. M. Comparison of historical and contemporary isolates of *Senecavirus A*. **Veterinary Microbiology**, [s. l.], v. 253, p. 108946, 2021.

BUCKLEY, A.; CRAWFORD, L.; HOFFMAN, K.; FALKENBERG, S. Experimental *Senecavirus A* Infection of Bovine Cell Lines and Colostrum-Deprived Calves. **Viruses**, [s. l.], v. 14, n. 12, p. 2809, 2022.

BUCKLEY, A.; LAGER, K. Efficacy of an inactivated *Senecavirus A* vaccine in weaned pigs and mature sows. **Vaccine**, [s. l.], v. 40, n. 12, p. 1747–1754, 2022.

CASERTA, L. C.; *et al.* Stability of *Senecavirus A* in animal feed ingredients and infection following consumption of contaminated feed. **Transboundary and Emerging Diseases**, [s. l.], v. 69, n. 1, p. 88–96, 2022.

CANTALUPO, P. G.; PIPAS, J. M. Detecting viral sequences in NGS data. **Current Opinion in Virology**, [s. l.], v. 39, p. 41–48, 2019.

CAO, L.; ZHANG, R.; LIU, T.; SUN, Z.; HU, M.; SUN, Y.; CHENG, L.; GUO, Y.; FU, S.; HU, J.; LI, X.; YU, C.; WANG, H.; CHEN, H.; LI, X.; FRY, E. E.; STUART, D. I.; QIAN, P.; LOU, Z.; RAO, Z. *Seneca Valley virus* attachment and uncoating mediated by its receptor anthrax toxin receptor 1. **Proceedings of the National Academy of Sciences**, [s. l.], v. 115, n. 51, p. 13087–13092, 2018.

CHEN, M.; CHEN, L.; WANG, J.; MOU, C.; CHEN, Z. Identification of a B-Cell Epitope in the VP3 Protein of *Senecavirus A*. **Viruses**, [s. l.], v. 13, n. 11, p. 2300, 2021.

CHEN, P. R.; ROWLAND, R. R. R.; STOIAN, A. M.; PETROVAN, V.; SHEAHAN, M.; GANTA, C.; CINO-OZUNA, G.; KIM, D. Y.; DUNLEAVEY, J. M.; WHITWORTH, K. M.; SAMUEL, M. S.; SPATE, L. D.; CECIL, R. F.; BENNE, J. A.; YAN, X.; FANG, Y.; CROIX, B. St.; LECHTENBERG, K.; WELLS, K. D.; PRATHER, R. S. Disruption of anthrax toxin receptor 1 in pigs leads to a rare disease phenotype and protection from *Senecavirus A* infection. **Scientific Reports**, [s. l.], v. 12, n. 1, p. 5009, 2022.

CHEN, X.; ZARO, J. L.; SHEN, W.-C. Fusion protein linkers: Property, design and functionality. **Advanced Drug Delivery Reviews**, [s. l.], v. 65, n. 10, p. 1357–1369, 2013.

CHEN, Y.; XU, Q.; TAN, C.; LI, X.; CHI, X.; CAI, B.; YU, Z.; MA, Y.; CHEN, J.-L. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on *Senecavirus A* evolution. **Microbial Pathogenesis**, [s. l.], v. 112, p. 313–319, 2017.

CHOUDHURY, S. M.; MA, X.; ZENG, Z.; LUO, Z.; LI, Y.; NIAN, X.; MA, Y.; SHI, Z.; SONG, R.; ZHU, Z.; CAO, W.; PEI, J.; ZHENG, H. *Senecavirus A* 3D Interacts with NLRP3 to Induce IL-1 β Production by Activating NF- κ B and Ion Channel Signals. **Microbiology Spectrum**, [s. l.], v. 10, n. 2, p. e02097-21, 2022.

DALL AGNOL, A. M.; OTONEL, R. A. A.; LEME, R. A.; ALFIERI, A. A.; ALFIERI, A. F. A TaqMan-based qRT-PCR assay for *Senecavirus A* detection in tissue samples of neonatal piglets. **Molecular and Cellular Probes**, [s. l.], v. 33, p. 28–31, 2017.

DHANDA, S. K.; VIR, P.; RAGHAVA, G. P. Designing of interferon-gamma inducing MHC class-II binders. **Biology Direct**, [s. l.], v. 8, n. 1, p. 30, 2013.

DIMITROV, I.; BANGOV, I.; FLOWER, D. R.; DOYTCHINOVA, I. AllerTOP v.2—a server for in silico prediction of allergens. **Journal of Molecular Modeling**, [s. l.], v. 20, n. 6, p. 2278, 2014.

DOYTCHINOVA, I. A.; FLOWER, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. **BMC Bioinformatics**, [s. l.], v. 8, n. 1, p. 4, 2007.

DU, L.; LI, B.; PANG, F.; YU, Z.; XU, X.; FAN, B.; TAN, Y.; HE, K.; HUANG, K. Porcine GPX1 enhances GP5-based DNA vaccination against porcine reproductive and respiratory syndrome virus. **Veterinary Immunology and Immunopathology**, [s. l.], v. 183, p. 31–39, 2017.

DVORAK, C. M. T.; AKKUTAY-YOLDAR, Z.; STONE, S. R.; TOUSIGNANT, S. J. P.; VANNUCCI, F. A.; MURTAUGH, M. P. An indirect enzyme-linked immunosorbent assay for the identification of antibodies to *Senecavirus A* in swine. **BMC Veterinary Research**, [s. l.], v. 13, n. 1, p. 50, 2017.

EJIGU, G. F.; JUNG, J. Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. **Biology**, [s. l.], v. 9, n. 9, p. 295, 2020.

FAN, H.; ZHU, H.; LI, S.; SHI, M.; ZHOU, E.; WANG, X.; JIANG, P.; BAI, J. Identification of linear B cell epitopes on VP1 and VP2 proteins of *Senecavirus A* (SVA) using monoclonal antibodies. **Veterinary Microbiology**, [s. l.], v. 247, p. 108753, 2020.

FERONATO, C.; LEME, R. A.; DINIZ, J. A.; AGNOL, A. M. D.; ALFIERI, A. F.; ALFIERI, A. A. Development and evaluation of a nested-PCR assay for *Senecavirus A* diagnosis. **Tropical Animal Health and Production**, [s. l.], v. 50, n. 2, p. 337–344, 2018.

FIETTO, Juliana Lopes Rangel; MACIEL, Talles Eduardo Ferreira. Sequenciando genomas. In: MOREIRA, Leandro Marcio. Ciências genômicas: fundamentos e aplicações. Ribeirão Preto: Sociedade Brasileira de Genética, 2015.

FOWLER, V. L.; RANSBURGH, R. H.; POULSEN, E. G.; WADSWORTH, J.; KING, D. P.; MIOULET, V.; KNOWLES, N. J.; WILLIAMSON, S.; LIU, X.; ANDERSON, G. A.; FANG, Y.; BAI, J. Development of a novel real-time RT-PCR assay to detect *Seneca Valley virus-1* associated with emerging cases of vesicular disease in pigs. **Journal of Virological Methods**, [s. l.], v. 239, p. 34–37, 2017.

GAO, H.; CHEN, Y.; XU, X.; XU, Z.; XU, S.; XING, J.; LIU, J.; ZHA, Y.; SUN, Y.; ZHANG, G. Comprehensive phylogeographic and phylodynamic analyses of global *Senecavirus A*. **Frontiers in Microbiology**, [s. l.], v. 13, p. 980862, 2022.

GASTEIGER, E.; HOOGLAND, C.; GATTIKER, A.; DUVAUD, S.; WILKINS, M. R.; APPEL, R. D.; BAIROCH, A. Protein Identification and Analysis Tools on the ExPASy Server. Em: WALKER, J. M. (Ed.). **The Proteomics Protocols Handbook**. Totowa, NJ: Humana Press, 2005. p. 571–607.

GIMENEZ-LIROLA, L. G.; RADEMACHER, C.; LINHARES, D.; HARMON, K.; ROTOLO, M.; SUN, Y.; BAUM, D. H.; ZIMMERMAN, J.; PIÑEYRO, P. Serological and Molecular Detection of *Senecavirus A* Associated with an Outbreak of Swine Idiopathic Vesicular Disease

and Neonatal Mortality. **Journal of Clinical Microbiology**, [s. l.], v. 54, n. 8, p. 2082–2089, 2016.

GOOLIA, M.; VANNUCCI, F.; YANG, M.; PATNAYAK, D.; BABIUK, S.; NFON, C. K. Validation of a competitive ELISA and a virus neutralization test for the detection and confirmation of antibodies to *Senecavirus A* in swine sera. **Journal of Veterinary Diagnostic Investigation**, [s. l.], v. 29, n. 2, p. 250–253, 2017.

GROTE, A.; HILLER, K.; SCHEER, M.; MUNCH, R.; NORTEMANN, B.; HEMPEL, D. C.; JAHN, D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. **Nucleic Acids Research**, [s. l.], v. 33, n. Web Server, p. W526–W531, 2005.

GUPTA, S.; KAPOOR, P.; CHAUDHARY, K.; GAUTAM, A.; KUMAR, R.; CONSORTIUM, O. S. D. D.; RAGHAVA, G. P. S. *In Silico* Approach for Predicting Toxicity of Peptides and Proteins. **PLOS ONE**, [s. l.], v. 8, n. 9, p. e73957, 2013.

HALES, L. M.; KNOWLES, N. J.; REDDY, P. S.; XU, L.; HAY, C.; HALLENBECK, P. L. Complete genome sequence analysis of *Seneca Valley virus-001*, a novel oncolytic picornavirus. **Journal of General Virology**, [s. l.], v. 89, n. 5, p. 1265–1275, 2008.

HEBDITCH, M.; CARBALLO-AMADOR, M. A.; CHARONIS, S.; CURTIS, R.; WARWICKER, J. Protein–Sol: a web tool for predicting protein solubility from sequence. **Bioinformatics**, [s. l.], v. 33, n. 19, p. 3098–3100, 2017.

HEINSON, A. I.; WOELK, C. H.; NEWELL, M.-L. The promise of reverse vaccinology. **International Health**, [s. l.], v. 7, n. 2, p. 85–89, 2015.

HEO, L.; PARK, H.; SEOK, C. GalaxyRefine: protein structure refinement driven by side-chain repacking. **Nucleic Acids Research**, [s. l.], v. 41, n. W1, p. W384–W388, 2013.

HOLE, K.; AHMADPOUR, F.; KRISHNAN, J.; STANSFIELD, C.; COPPS, J.; NFON, C. Efficacy of accelerated hydrogen peroxide[®] disinfectant on foot-and-mouth disease virus, swine vesicular disease virus and *Senecavirus A*. **Journal of Applied Microbiology**, [s. l.], v. 122, n. 3, p. 634–639, 2017.

HOU, L.; TONG, X.; PAN, Y.; SHI, R.; LIU, C.; GUO, J.; SHI, Y.; YANG, X.; WANG, Y.; FENG, X.; ZHOU, J.; LIU, J. *Seneca Valley Virus* Enters PK-15 Cells via Caveolae-Mediated Endocytosis and Macropinocytosis Dependent on Low-pH, Dynamin, Rab5, and Rab7. **Journal of Virology**, [s. l.], v. 96, n. 24, p. e01446-22, 2022.

HOUSTON, E.; GIMÉNEZ-LIROLA, L. G.; MAGTOTO, R.; MORA-DÍAZ, J. C.; BAUM, D.; PIÑEYRO, P. E. Seroprevalence of *Senecavirus A* in sows and grower-finisher pigs in major swine producing-states in the United States. **Preventive Veterinary Medicine**, [s. l.], v. 165, p. 1–7, 2019.

HOUSTON, E.; TEMEEYASEN, G.; PIÑEYRO, P. E. Comprehensive review on immunopathogenesis, diagnostic and epidemiology of *Senecavirus A*. **Virus Research**, [s. l.], v. 286, p. 198038, 2020.

HOWE, K.; CHOW, W.; COLLINS, J.; PELAN, S.; POINTON, D.-L.; SIMS, Y.; TORRANCE, J.; TRACEY, A.; WOOD, J. Significantly improving the quality of genome assemblies through curation. **GigaScience**, [s. l.], v. 10, n. 1, p. giaa153, 2021.

HU, T.; CHITNIS, N.; MONOS, D.; DINH, A. Next-generation sequencing technologies: An overview. **Human Immunology**, [s. l.], v. 82, n. 11, p. 801–811, 2021.

ICTV - International Committee on Taxonomy of Viruses - *Senecavirus A*. Disponível em: <https://ictv.global/taxonomy/taxondetails?taxnode_id=202202061&taxon_name=Senecavirus%20A>. Acesso em 22/12/2023.

INATA. Bula da vacina Senevac. [s.d.], a. Disponível em: <<https://inata.com.br/wp-content/uploads/2023/02/Bula-Vacina-Senevac-Senecavirus-Inata.pdf>>. Acesso em: 28/12/2023.

JIA, M.; SUN, M.; TANG, Y.-D.; ZHANG, Y.-Y.; WANG, H.; CAI, X.; MENG, F. *Senecavirus A* Entry Into Host Cells Is Dependent on the Cholesterol-Mediated Endocytic Pathway. **Frontiers in Veterinary Science**, [s. l.], v. 9, p. 840655, 2022.

JOSHI, L. R.; FERNANDES, M. H. V.; CLEMENT, T.; LAWSON, S.; PILLATZKI, A.; RESENDE, T. P.; VANNUCCI, F. A.; KUTISH, G. F.; NELSON, E. A.; DIEL, D. G. Pathogenesis of *Senecavirus A* infection in finishing pigs. **Journal of General Virology**, [s. l.], v. 97, n. 12, p. 3267–3279, 2016. a.

JOSHI, L. R.; MOHR, K. A.; CLEMENT, T.; HAIN, K. S.; MYERS, B.; YAROS, J.; NELSON, E. A.; CHRISTOPHER-HENNINGS, J.; GAVA, D.; SCHAEFER, R.; CARON, L.; DEE, S.; DIEL, D. G. Detection of the Emerging Picornavirus *Senecavirus A* in Pigs, Mice, and Houseflies. **Journal of Clinical Microbiology**, [s. l.], v. 54, n. 6, p. 1536–1545, 2016. b.

JOSHI, L. R.; MOHR, K. A.; GAVA, D.; KUTISH, G.; BUYASSE, A. S.; VANNUCCI, F. A.; PIÑEYRO, P. E.; CROSSLEY, B. M.; SCHILTZ, J. J.; JENKINS-MOORE, M.; KOSTER, L.; TELL, R.; SCHAEFER, R.; MARTHALER, D.; DIEL, D. G. Genetic diversity and evolution of the emerging picornavirus *Senecavirus A*. **Journal of General Virology**, [s. l.], v. 101, n. 2, p. 175–187, 2020.

JUMPER, J.; EVANS, R.; PRITZEL, A.; GREEN, T.; FIGURNOV, M.; RONNEBERGER, O.; TUNYASUVUNAKOOL, K.; BATES, R.; ŽÍDEK, A.; POTAPENKO, A.; BRIDGLAND, A.; MEYER, C.; KOHL, S. A. A.; BALLARD, A. J.; COWIE, A.; ROMERA-PAREDES, B.; NIKOLOV, S.; JAIN, R.; ADLER, J.; BACK, T.; PETERSEN, S.; REIMAN, D.; CLANCY, E.; ZIELINSKI, M.; STEINEGGER, M.; PACHOLSKA, M.; BERGHAMMER, T.; BODENSTEIN, S.; SILVER, D.; VINYALS, O.; SENIOR, A. W.; KAVUKCUOGLU, K.; KOHLI, P.; HASSABIS, D. Highly accurate protein structure prediction with AlphaFold. **Nature**, [s. l.], v. 596, n. 7873, p. 583–589, 2021.

KREMER, Frederico Schmitt. Pós-montagem de Genomas (parte II): scaffolding. 2020. Disponível em: <<https://medium.com/omixdata/pós-montagem-de-genomas-parte-ii-scaffolding-7c78a218dfca>>. Acesso em: 19/01/2024.

KNOWLES, N.J.; *et al.* Epidemiology of *Seneca Valley virus*: Identification and characterization of isolates from pigs in the United States. In: Proceedings of the Abstracts of the Northern Lights EUROPIC 2006—XIV Meeting of the European Study Group on the Molecular Biology of Picornaviruses, Abstract G2. Disponível em: <https://www.researchgate.net/publication/267416696_Epidemiology_of_Seneca_Valley_Virus_Identification_and_Characterization_of_Isolates_from_Pigs_in_the_United_States>. Acesso em: 15/12/2023.

LAGUARDIA-NASCIMENTO, M.; GASPARINI, M. R.; SALES, É. B.; RIVETTI, A. V.; SOUSA, N. M.; OLIVEIRA, A. M.; CAMARGOS, M. F.; PINHEIRO DE OLIVEIRA, T. F.; GONÇALVES, J. P. M.; MADUREIRA, M. C.; RIBEIRO, D. P.; MARCONDES, I. V.; BARBOSA-STANCIOLI, E. F.; FONSECA, A. A. Molecular epidemiology of *Senecavirus A* associated with vesicular disease in pigs in Brazil. **The Veterinary Journal**, [s. l.], v. 216, p. 207–209, 2016.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, [s. l.], v. 9, n. 4, p. 357–359, 2012.

LARSEN, J.; LUND, O.; NIELSEN, M. Improved method for predicting linear B-cell epitopes. **Immunome Research**, [s. l.], v. 2, n. 1, p. 2, 2006.

LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, [s. l.], v. 26, n. 2, p. 283–291, 1993.

LEME, R. A.; OLIVEIRA, T. E. S.; ALFIERI, A. F.; HEADLEY, S. A.; ALFIERI, A. A. Pathological, Immunohistochemical and Molecular Findings Associated with *Senecavirus A*-Induced Lesions in Neonatal Piglets. **Journal of Comparative Pathology**, [s. l.], v. 155, n. 2–3, p. 145–155, 2016.

LEME, R. A.; ZOTTI, E.; ALCÂNTARA, B. K.; OLIVEIRA, M. V.; FREITAS, L. A.; ALFIERI, A. F.; ALFIERI, A. A. *Senecavirus A*: An Emerging Vesicular Infection in Brazilian Pig Herds. **Transboundary and Emerging Diseases**, [s. l.], v. 62, n. 6, p. 603–611, 2015.

LEME, R.; ALFIERI, A.; ALFIERI, A. Update on Senecavirus Infection in Pigs. **Viruses**, [s. l.], v. 9, n. 7, p. 170, 2017.

LI, J.; LIANG, W.; XU, S.; SHI, J.; ZHOU, X.; LIU, B.; YU, L.; XIONG, J.; SI, G.; HE, D. Rapid and sensitive detection of *Senecavirus A* by reverse transcription loop-mediated isothermal amplification combined with a lateral flow dipstick method. **PLOS ONE**, [s. l.], v. 14, n. 5, p. e0216245, 2019.

LI, N.; QIAO, Q.; GUO, H.; WANG, B.; HUANG, Q.; WANG, Z.; LI, Y.; ZHAO, J. Evaluation of immunogenicity and protective efficacy of a novel *Senecavirus A* strain-based inactivated vaccine in mice. **Research in Veterinary Science**, [s. l.], v. 142, p. 133–140, 2022.

LI, Y.; ZHANG, Y.; LIAO, Y.; SUN, Y.; RUAN, Y.; LIU, C.; ZHANG, M.; LI, F.; LI, X.; FAN, S.; YI, L.; DING, H.; ZHAO, M.; FAN, J.; CHEN, J. Preliminary Evaluation of Protective Efficacy of Inactivated *Senecavirus A* on Pigs. **Life**, [s. l.], v. 11, n. 2, p. 157, 2021.

LINHARES, D. *et al.* Seneca-Associated Diseases, Clinical Presentation And Epidemiological Distribution. Apresentação. Allen D. Leman Swine Conference, 2015. Disponível em: <<https://www.slideshare.net/trufflemedia/dr-daniel-linhares-senecaassociated-diseases-clinical-presentation-and-epidemiological-distribution>>. Acesso em: 15/12/2023.

LIU, F.; WANG, Q.; HUANG, Y.; WANG, N.; SHAN, H. A 5-Year Review of *Senecavirus A* in China since Its Emergence in 2015. **Frontiers in Veterinary Science**, [s. l.], v. 7, p. 567792, 2020.

LIU, H.; LI, K.; CHEN, W.; YANG, F.; CAO, W.; ZHANG, K.; LI, P.; TANG, L.; ZHU, Z.; ZHENG, H. *Senecavirus A* 2B protein suppresses type I interferon production by inducing the degradation of MAVS. **Molecular Immunology**, [s. l.], v. 142, p. 11–21, 2022. a.

LIU, W.; LI, X.; ZHANG, H.; HAO, G.; SHANG, X.; WANG, H.; CHEN, H.; QIAN, P. Evaluation of Immunoreactivity and Protection Efficacy of *Seneca Valley Virus* Inactivated Vaccine in Finishing Pigs Based on Screening of Inactivated Agents and Adjuvants. **Vaccines**, [s. l.], v. 10, n. 4, p. 631, 2022. b.

LIU, Z.; *et al.* Intravenous injection of oncolytic picornavirus SVV-001 prolongs animal survival in a panel of primary tumor-based orthotopic xenograft mouse models of pediatric glioma. **Neuro-Oncology**, [s. l.], v. 15, n. 9, p. 1173–1185, 2013.

LO, Y.-T.; SHIH, T.-C.; PAI, T.-W.; HO, L.-P.; WU, J.-L.; CHOU, H.-Y. Conformational epitope matching and prediction based on protein surface spiral features. **BMC Genomics**, [s. l.], v. 22, n. S2, p. 116, 2021.

LUO, A.; QIAO, H.; ZHANG, Y.; SHI, W.; HO, S. Y.; XU, W.; ZHANG, A.; ZHU, C. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. **BMC Evolutionary Biology**, [s. l.], v. 10, n. 1, p. 242, 2010.

MA, X.; HUANG, J.; LI, K.; DING, K.; FU, Y.; ZHANG, J.; ZHAO, Z.; LI, P.; BAI, X.; LI, D.; LIU, X.; ZENG, Q.; LIU, Z.; SUN, P.; LU, Z. Development and Evaluation of a Competitive Enzyme-Linked Immunosorbent Assay Based on Swine Monoclonal Antibodies for Detecting Neutralizing Antibodies against *Senecavirus A*. **Microbiology Spectrum**, [s. l.], v. 11, n. 3, p. e04599-22, 2023.

MA, Z.; LV, J.; ZHANG, Z.; PAN, L. Development of an indirect ELISA using a novel linear epitope at the C-terminal region of the VP2 protein to specifically detect antibodies against *Senecavirus A*. **Virology Journal**, [s. l.], v. 19, n. 1, p. 204, 2022.

MAGGIOLI, M. F.; FERNANDES, M. H. V.; JOSHI, L. R.; SHARMA, B.; TWEET, M. M.; NOLL, J. C. G.; BAUERMANN, F. V.; DIEL, D. G. Persistent Infection and Transmission of *Senecavirus A* from Carrier Sows to Contact Piglets. **Journal of Virology**, [s. l.], v. 93, n. 21, p. e00819-19, 2019.

MAGGIOLI, M. F.; LAWSON, S.; DE LIMA, M.; JOSHI, L. R.; FACCIN, T. C.; BAUERMANN, F. V.; DIEL, D. G. Adaptive Immune Responses following *Senecavirus A* Infection in Pigs. **Journal of Virology**, [s. l.], v. 92, n. 3, p. e01717-17, 2018.

MARDIS, E. R. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, AnnualReviews,v.9,p.387–402,2008.

MCDONNELL, E.; STRASSER, K.; TSANG, A. Manual Gene Curation and Functional Annotation. Em: DE VRIES, R. P.; TSANG, A.; GRIGORIEV, I. V. (Eds.). **Fungal Genomics. Methods in Molecular Biology**New York, NY: Springer New York, 2018. v. 1775p. 185–208.

MCGUFFIN, L. J.; BRYSON, K.; JONES, D. T. The PSIPRED protein structure prediction server. **Bioinformatics**, [s. l.], v. 16, n. 4, p. 404–405, 2000.

MILES, L. A.; BURGA, L. N.; GARDNER, E. E.; BOSTINA, M.; POIRIER, J. T.; RUDIN, C. M. Anthrax toxin receptor 1 is the cellular receptor for *Seneca Valley virus*. **Journal of Clinical Investigation**, [s. l.], v. 127, n. 8, p. 2957–2967, 2017.

MIRDITA, M.; SCHÜTZE, K.; MORIWAKI, Y.; HEO, L.; OVCHINNIKOV, S.; STEINEGGER, M. ColabFold: making protein folding accessible to all. **Nature Methods**, [s. l.], v. 19, n. 6, p. 679–682, 2022.

MU, S.; SUN, S.; DONG, H.; BAI, M.; ZHANG, Y.; TENG, Z.; REN, M.; YIN, S.; GUO, H. Potent Protective Immune Responses to Senecavirus Induced by Virus-Like Particle Vaccine in Pigs. **Vaccines**, [s. l.], v. 8, n. 3, p. 532, 2020.

NAVARRO-LOPEZ, R.; *et al.* First detection and genetic characterization of *Senecavirus A* in pigs from Mexico. **Journal of Swine Health and Production**, [s. l.], v. 31, n. 6, p. 289–294, 2023.

NEATH, A. A.; CAVANAUGH, J. E. The Bayesian information criterion: background, derivation, and applications. **WIREs Computational Statistics**, [s. l.], v. 4, n. 2, p. 199–203, 2012.

OLSON, R. D.; *et al.* Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. **Nucleic Acids Research**, [s. l.], v. 51, n. D1, p. D678–D689, 2023.

PANJWANI, A.; STRAUSS, M.; GOLD, S.; WENHAM, H.; JACKSON, T.; CHOU, J. J.; ROWLANDS, D. J.; STONEHOUSE, N. J.; HOGLE, J. M.; TUTHILL, T. J. Capsid Protein VP4 of Human Rhinovirus Induces Membrane Permeability by the Formation of a Size-Selective Multimeric Pore. **PLoS Pathogens**, [s. l.], v. 10, n. 8, p. e1004294, 2014.

PASMA, T.; DAVIDSON, S.; SHAW, S. L. Idiopathic vesicular disease in swine in Manitoba. **Can. Vet. J.**, [s. l.], v. 49, p. 84–85, 2008.

PERVEZ, M. T.; HASNAIN, M. J. U.; ABBAS, S. H.; MOUSTAFA, M. F.; ASLAM, N.; SHAH, S. S. M. A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. **BioMed Research International**, [s. l.], v. 2022, p. 1–12, 2022.

PINHEIRO-DE-OLIVEIRA, T. F.; FONSECA-JÚNIOR, A. A.; CAMARGOS, M. F.; LAGUARDIA-NASCIMENTO, M.; GIANNATTASIO-FERRAZ, S.; COTTORRELLA, A. C. P.; DE OLIVEIRA, A. M.; GÓES-NETO, A.; BARBOSA-STANCIOLI, E. F. Reverse transcriptase droplet digital PCR to identify the emerging vesicular virus *Senecavirus A* in biological samples. **Transboundary and Emerging Diseases**, [s. l.], v. 66, n. 3, p. 1360–1369, 2019.

PLOTKIN, S.; ROBINSON, J. M.; CUNNINGHAM, G.; IQBAL, R.; LARSEN, S. The complexity and cost of vaccine manufacturing – An overview. **Vaccine**, [s. l.], v. 35, n. 33, p. 4064–4071, 2017.

PONOMARENKO, J.; BUI, H.-H.; LI, W.; FUSSEDER, N.; BOURNE, P. E.; SETTE, A.; PETERS, B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. **BMC Bioinformatics**, [s. l.], v. 9, n. 1, p. 514, 2008.

PREIS, G. M. *Senecavirus A* in pigs: Epidemiology, Transmission, and Diagnosis. 2022. University of Minnesota, Saint Paul, Minnesota, USA, 2022.

QADEER, S.; KHAN, M. S.; JOYIA, F. A.; ZIA, M. A. Immunogenic profiling and designing of a novel vaccine from capsid proteins of FMDV serotype Asia-1 through reverse vaccinology. **Infection, Genetics and Evolution**, [s. l.], v. 93, p. 104925, 2021.

QUER, J.; *et al.* Next-Generation Sequencing for Confronting Virus Pandemics. **Viruses**, [s. l.], v. 14, n. 3, p. 600, 2022.

RAPIN, N.; LUND, O.; BERNASCHI, M.; CASTIGLIONE, F. Computational Immunology Meets Bioinformatics: The Use of Prediction Tools for Molecular Binding in the Simulation of the Immune System. **PLoS ONE**, [s. l.], v. 5, n. 4, p. e9862, 2010.

REED, L.J., MUENCH, H. A simple method of estimating fifty percent endpoints. *Am. J. Hyg.* v.27, p.493–49. 1938.

RESENDE, T. P.; MARTHALER, D. G.; VANNUCCI, F. A. A novel RNA-based in situ hybridization to detect *Seneca Valley virus* in neonatal piglets and sows affected with vesicular disease. **PLOS ONE**, [s. l.], v. 12, n. 4, p. e0173190, 2017.

REYNISSON, B.; ALVAREZ, B.; PAUL, S.; PETERS, B.; NIELSEN, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. **Nucleic Acids Research**, [s. l.], v. 48, n. W1, p. W449–W454, 2020.

RIEGER, J. S. G.; *et al.* Standardization of ELISA with *Senecavirus A* recombinant VP2 protein and its use in swine herds in Brazil. **Genetics and Molecular Research**, [s. l.], v. 22, n. 1, 2023. Disponível em: <http://www.funpecrp.com.br/gmr/articles/year2023/vol22-1/pdf/gmr19118_-_standardization-elisa-senecavirus-recombinant-vp2-protein-and-its-use-swine-herds-brazil.pdf>. Acesso em: 15 dez. 2023.

ROCHA, B. Desenvolvimento da técnica de imunoperoxidase em monocamada de células para detecção de anticorpos contra o *Senecavirus A* em amostras de soro de suínos. 2022. Dissertação de mestrado - Universidade Federal de Minas Gerais, Belo Horizonte, 2022.

ROTHBERG, Jonathan M.; LEAMON, John H. The development and impact of 454 sequencing. **Nature Biotechnology**, [s.l.], v. 26, n. 10, p. 1117-1124, 2008.

RU, Y.; HAO, R.; WU, C.; LI, Y.; LU, B.; LIU, H.; TIAN, H.; LI, D.; SHI, Z.; LUO, J.; MA, K.; ZHANG, G.; LIU, X.; ZHENG, H. Identification of Potential Novel B-Cell Epitopes of Capsid Protein VP2 in *Senecavirus A*. **Microbiology Spectrum**, [s. l.], v. 11, n. 4, p. e04472-22, 2023.

RUSTON, C.; ZHANG, J.; SCOTT, J.; ZHANG, M.; GRAHAM, K.; LINHARES, D.; BREUER, M.; KARRIKER, L.; HOLTkamp, D. Efficacy of ultraviolet C exposure for inactivating *Senecavirus A* on experimentally contaminated surfaces commonly found on swine farms. **Veterinary Microbiology**, [s. l.], v. 256, p. 109040, 2021.

SAENG-CHUTO, K.; *et al.* The first detection of *Senecavirus A* in pigs in Thailand, 2016. *Transboundary and Emerging Diseases*, 65, 285–288, 2018.

SAHA, S.; RAGHAVA, G. P. S. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. **Proteins: Structure, Function, and Bioinformatics**, [s. l.], v. 65, n. 1, p. 40–48, 2006.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 74, n. 12, p. 5463–5467, 1977.

SANTOS, B. *et al.* Nanotechnology meets immunology towards a rapid diagnosis solution: the COVID-19 outbreak challenge. **RSC Advances**, [s. l.], v. 12, n. 49, p. 31711–31728, 2022.

SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic datasets. **Bioinformatics**, [s. l.], v. 27, n. 6, p. 863–864, 2011.

SEGALÉS, J.; BARCELLOS, D.; ALFIERI, A.; BURROUGH, E.; MARTHALER, D. *Senecavirus A*: An Emerging Pathogen Causing Vesicular Disease and Mortality in Pigs? **Veterinary Pathology**, [s. l.], v. 54, n. 1, p. 11–21, 2017.

SHAMS, M.; NOURMOHAMMADI, H.; ASGHARI, A.; BASATI, G.; MAJIDIANI, H.; NASERIFAR, R.; IRANNEJAD, H. Construction of a multi-epitope protein for human *Toxocara canis* detection: Immunoinformatics approach multi-epitope construct for *T. canis* serodiagnosis. **Informatics in Medicine Unlocked**, [s. l.], v. 26, p. 100732, 2021.

SHARMA, B.; FERNANDES, M. H. V.; DE LIMA, M.; JOSHI, L. R.; LAWSON, S.; DIEL, D. G. A Novel Live Attenuated Vaccine Candidate Protects Against Heterologous *Senecavirus A* Challenge. **Frontiers in Immunology**, [s. l.], v. 10, p. 2660, 2019.

SHAWAN, M. M. A. K.; SHARMA, A. R.; HALDER, S. K.; ARIAN, T. A.; SHUVO, Md. N.; SARKER, S. R.; HASAN, Md. A. Advances in Computational and Bioinformatics Tools and Databases for Designing and Developing a Multi-Epitope-Based Peptide Vaccine. **International Journal of Peptide Research and Therapeutics**, [s. l.], v. 29, n. 4, p. 60, 2023.

SINGH, A.; MOR, S.; ABOUBAKR, H.; VANNUCCI, F.; PATNAYAK, D.; GOYAL, S. Efficacy of three disinfectants against *Senecavirus A* on five surfaces and at two temperatures. **Journal of Swine Health and Production**, [s. l.], v. 25, n. 2, p. 64–68, 2017.

SINGH, K.; CORNER, S.; CLARK, S. G.; SCHERBA, G.; FREDRICKSON, R. *Seneca Valley Virus* and Vesicular Lesions in a Pig with Idiopathic Vesicular Disease. **Journal of Veterinary Science & Technology**, [s. l.], v. 03, n. 06, 2012. Disponível em: <<https://www.omicsonline.org/seneca-valley-virus-and-vesicular-lesions-in-a-pig-with-idiopathic-vesicular-disease-2157-7579.1000123.php?aid=9061>>. Acesso em: 14 dez. 2023.

SNAPGENE. Software for everyday molecular biology. [s.d.]. Disponível em: <<https://www.snapgene.com/>>. Acesso em: 04/01/2023.

SONG, J.; HOU, L.; QUAN, R.; WANG, D.; JIANG, H.; LIU, J. Synergetic Contributions of Viral VP1, VP3, and 3C to Activation of the AKT-AMPK-MAPK-MTOR Signaling Pathway for *Seneca Valley Virus*-Induced Autophagy. **Journal of Virology**, [s. l.], v. 96, n. 2, p. e01550-21, 2022.

STAATS, C. C.; MORAIS, G. L.; MARGIS, R. Projetos genoma. In: VERLI, Hugo. *Bioinformática: da biologia à flexibilidade molecular*. 1. ed. São Paulo: SBBq, 2014.

STRAUSS, M.; JAYAWARDENA, N.; SUN, E.; EASINGWOOD, R. A.; BURGA, L. N.; BOSTINA, M. Cryo-Electron Microscopy Structure of Seneca Valley Virus Procapsid. **Journal of Virology**, [s. l.], v. 92, n. 6, p. e01927-17, 2018.

STUROS, M. J.; MURRAY, D.; JOHNSON, L.; PREIS, G.; CORZO, C. A.; ROSSOW, S.; VANNUCCI, F. A. Persistence and shedding of *Senecavirus A* in naturally infected boars. **Journal of Veterinary Diagnostic Investigation**, [s. l.], v. 34, n. 3, p. 474–481, 2022.

SUN, D.; VANNUCCI, F.; KNUTSON, T. P.; CORZO, C.; MARTHALER, D. G. Emergence and whole-genome sequence of *Senecavirus A* in Colombia. **Transboundary and Emerging Diseases**, [s. l.], v. 64, n. 5, p. 1346–1349, 2017.

SUN, Z.; LAWSON, S.; LANGENHORST, R.; MCCORMICK, K. L.; BRUNICK, C.; OPRIESSNIG, T.; BAKER, R.; YOON, K.-J.; ZHANG, W.; HUBER, V. C.; FANG, Y. Construction and immunogenicity evaluation of an epitope-based antigen against swine influenza A virus using *Escherichia coli* heat-labile toxin B subunit as a carrier–adjuvant. **Veterinary Microbiology**, [s. l.], v. 164, n. 3–4, p. 229–238, 2013.

TAO, Q.; ZHU, L.; XU, L.; YANG, Y.; ZHANG, Y.; LIU, Z.; XU, T.; WEN, J.; DENG, L.; ZHOU, Y.; XU, Z. The Construction and Immunogenicity Analyses of a Recombinant Pseudorabies Virus with *Senecavirus A* VP2 Protein Coexpression. **Microbiology Spectrum**, [s. l.], v. 11, n. 2, p. e05229-22, 2023.

THE UNIPROT CONSORTIUM et al. UniProt: the Universal Protein Knowledgebase in 2023. **Nucleic Acids Research**, [s. l.], v. 51, n. D1, p. D523–D531, 2023.

TOMAZIC, M. L.; RODRIGUEZ, A. E.; LOMBARDELLI, J.; POKLEPOVICH, T.; GARRO, C.; GALARZA, R.; TIRANTI, K.; FLORIN-CHRISTENSEN, M.; SCHNITTGER, L. Identification of novel vaccine candidates against cryptosporidiosis of neonatal bovines by reverse vaccinology. **Veterinary Parasitology**, [s. l.], v. 264, p. 74–78, 2018.

TORCHALA, M.; MOAL, I. H.; CHALEIL, R. A. G.; FERNANDEZ-RECIO, J.; BATES, P. A. SwarmDock: a server for flexible protein–protein docking. **Bioinformatics**, [s. l.], v. 29, n. 6, p. 807–809, 2013.

TUNYASUVUNAKOOL, K.; *et al.* Highly accurate protein structure prediction for the human proteome. **Nature**, [s. l.], v. 596, n. 7873, p. 590–596, 2021.

TURNER, J. H.; PAIM, W. P.; MAGGIOLI, M. F.; PETER, C. M.; MIKNIS, R.; TALLEY, J.; BAUERMAN, F. V. Prolonged Viability of *Senecavirus A* in Exposed House Flies (*Musca domestica*). **Viruses**, [s. l.], v. 14, n. 1, p. 127, 2022.

VALE, D. L.; *et al.* Serodiagnosis of canine leishmaniasis using a novel recombinant chimeric protein constructed with distinct B-cell epitopes from antigenic *Leishmania infantum* proteins. **Veterinary Parasitology**, [s. l.], v. 296, p. 109513, 2021.

VANNUCCI, F. A.; LINHARES, D. C. L.; BARCELLOS, D. E. S. N.; LAM, H. C.; COLLINS, J.; MARTHALER, D. Identification and Complete Genome of *Seneca Valley Virus* in Vesicular

Fluid and Sera of Pigs Affected with Idiopathic Vesicular Disease, Brazil. **Transboundary and Emerging Diseases**, [s. l.], v. 62, n. 6, p. 589–593, 2015.

VENKATARAMAN, S.; REDDY, S. P.; LOO, J.; IDAMAKANTI, N.; HALLENBECK, P. L.; REDDY, V. S. Crystallization and preliminary X-ray diffraction studies of *Seneca Valley Virus-001*, a new member of the *Picornaviridae* family. **Acta Crystallographica Section F Structural Biology and Crystallization Communications**, [s. l.], v. 64, n. 4, p. 293–296, 2008.

VIEIRA, M. V.; YASUMITSU, C. Y.; DALL AGNOL, A. M.; LEME, R. A.; ALFIERI, A. F.; ALFIERI, A. A. The third wave of *Seneca Valley virus* outbreaks in pig herds in southern Brazil. **Brazilian Journal of Microbiology**, [s. l.], v. 53, n. 3, p. 1701–1706, 2022.

VILELA RODRIGUES, T. C.; JAISWAL, A. K.; LEMES, M. R.; DA SILVA, M. V.; SALES-CAMPOS, H.; ALCÂNTARA, L. C. J.; TOSTA, S. F. de O.; KATO, R. B.; ALZHRANI, K. J.; BARH, D.; AZEVEDO, V. A. de C.; TIWARI, S.; SOARES, S. de C. An immunoinformatics-based designed multi-epitope candidate vaccine (mpme-VAC/STV-1) against *Mycoplasma pneumoniae*. **Computers in Biology and Medicine**, [s. l.], v. 142, p. 105194, 2022.

VISHNU, U. S.; SANKARASUBRAMANIAN, J.; GUNASEKARAN, P.; RAJENDHRAN, J. Novel Vaccine Candidates against *Brucella melitensis* Identified through Reverse Vaccinology Approach. **OMICS: A Journal of Integrative Biology**, [s. l.], v. 19, n. 11, p. 722–729, 2015.

WANG, P.; MENG, F.; MOORE, B. M.; SHIU, S.-H. Impact of short-read sequencing on the misassembly of a plant genome. **BMC Genomics**, [s. l.], v. 22, n. 1, p. 99, 2021.

WANG, Q.; MENG, H.; GE, D.; SHAN, H.; GERI, L.; LIU, F. Structural and nonstructural proteins of *Senecavirus A*: Recent research advances, and lessons learned from those of other picornaviruses. **Virology**, [s. l.], v. 585, p. 155–163, 2023.

WU, H.; LI, C.; JI, Y.; MOU, C.; CHEN, Z.; ZHAO, J. The Evolution and Global Spatiotemporal Dynamics of *Senecavirus A*. **Microbiology Spectrum**, [s. l.], v. 10, n. 6, p. e02090-22, 2022.

XU, G.; XU, S.; SHI, X.; SHEN, C.; HAO, J.; YAN, M.; ZHANG, D.; ZHU, Z.; ZHANG, K.; ZHENG, H.; LIU, X. Intercellular transmission of *Seneca Valley virus* mediated by exosomes. **Veterinary Research**, [s. l.], v. 51, n. 1, p. 91, 2020.

XU, J.; MCPARTLON, M.; LI, J. Improved protein structure prediction by deep learning irrespective of co-evolution information. **Nature Machine Intelligence**, [s. l.], v. 3, n. 7, p. 601–609, 2021.

YAN, J.; GAO, Y.; LI, J.; LI, M.; GUO, C.; BAI, J.; JIANG, P. The Establishment and Application of Indirect 3AB-ELISA for the Detection of Antibodies against *Senecavirus A*. **Viruses**, [s. l.], v. 15, n. 4, p. 861, 2023.

YAN, Y.; ZHANG, D.; ZHOU, P.; LI, B.; HUANG, S.-Y. HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. **Nucleic Acids Research**, [s. l.], v. 45, n. W1, p. W365–W373, 2017.

YANG, F.; ZHU, Z.; CAO, W.; LIU, H.; ZHANG, K.; TIAN, H.; LIU, X.; ZHENG, H. Immunogenicity and protective efficacy of an inactivated cell culture-derived *Seneca Valley virus* vaccine in pigs. **Vaccine**, [s. l.], v. 36, n. 6, p. 841–846, 2018.

YANG, M.; VAN BRUGGEN, R.; XU, W. Generation and diagnostic application of monoclonal antibodies against *Seneca Valley virus*. **Journal of Veterinary Diagnostic Investigation**, [s. l.], v. 24, n. 1, p. 42–50, 2012.

YANG, W.; LI, D.; RU, Y.; BAI, J.; REN, J.; ZHANG, J.; LI, L.; LIU, X.; ZHENG, H. *Foot-and-Mouth Disease Virus* 3A Protein Causes Upregulation of Autophagy-Related Protein LRRC25 To Inhibit the G3BP1-Mediated RIG-Like Helicase-Signaling Pathway. **Journal of Virology**, [s. l.], v. 94, n. 8, p. e02086-19, 2020.

ZENG, F.; CONG, F.; LIU, X.; LIAN, Y.; WU, M.; XIAO, L.; YUAN, W.; HUANG, R.; MA, J.; GUO, P.; LUO, M. Development of a real time loop-mediated isothermal amplification method for detection of *Senecavirus A*. **Journal of Virological Methods**, [s. l.], v. 261, p. 98–103, 2018.

ZHANG, J.; *et al.* Development and evaluation of a real-time RT-PCR and a field-deployable RT-insulated isothermal PCR for the detection of *Seneca Valley virus*. **BMC Veterinary Research**, [s. l.], v. 15, n. 1, p. 168, 2019. a.

ZHANG, J.; *et al.* Full-Length Genome Sequences of *Senecavirus A* from Recent Idiopathic Vesicular Disease Outbreaks in U.S. Swine. **Genome Announcements**, [s. l.], v. 3, n. 6, p. e01270-15, 2015.

ZHANG, K.; XU, S.; SHI, X.; XU, G.; SHEN, C.; LIU, X.; ZHENG, H. Exosomes-mediated transmission of *foot-and-mouth disease virus* *in vivo* and *in vitro*. **Veterinary Microbiology**, [s. l.], v. 233, p. 164–173, 2019. b.

ZHANG, Z.; YAO, F.; LV, J.; DING, Y.; LIU, X.; ZHANG, L.; MA, Z.; ZHOU, P.; WANG, Y.; GUO, H.; PAN, L. Identification of B-cell epitopes on structural proteins VP1 and VP2 of *Senecavirus A* and development of a multi-epitope recombinant protein vaccine. **Virology**, [s. l.], v. 582, p. 48–56, 2023.

ZHAO, K.; ZHANG, S.; LIU, X.; GUO, X.; GUO, Z.; ZHANG, X.; YUAN, W. The game between host antiviral innate immunity and immune evasion strategies of *Senecavirus A* - A cell biological perspective. **Frontiers in Immunology**, [s. l.], v. 13, p. 1107173, 2022.

ZHAO, W.; SHER, X. Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. **PLOS Computational Biology**, [s. l.], v. 14, n. 11, p. e1006457, 2018.

ZHOU, J.; CHEN, J.; PENG, Y.; XIE, Y.; XIAO, Y. A Promising Tool in Serological Diagnosis: Current Research Progress of Antigenic Epitopes in Infectious Diseases. **Pathogens**, [s. l.], v. 11, n. 10, p. 1095, 2022.

ZHOU, X.; LIANG, W.-F.; SI, G.-B.; LI, J.-H.; CHEN, Z.-F.; CAI, W.-Y.; LV, D.-H.; WEN, X.-H.; ZHAI, Q.; ZHAI, S.-L.; LIAO, M.; HE, D.-S. Buffalo-Origin *Seneca Valley Virus* in China: First Report, Isolation, Genome Characterization, and Evolution Analysis. **Frontiers in Veterinary Science**, [s. l.], v. 8, p. 730701, 2021.

APÊNDICE I - Dados das 279 sequências de SVA disponíveis no NCBI *Virus*, com tamanho superior a 7000pb, utilizadas no presente trabalho (acesso em 30/10/2023).

Acesso	Origem	Tipo de amostra	Ano da coleta
KR063109	Brasil (GO)	Fluido vesicular (informado na publicação)	2015
KR063107	Brasil (MG)	Fluido vesicular (informado na publicação)	2015
KR063108	Brasil (MG)	Soro (informado na publicação)	2015
MF615508	Brasil (SC)	Cavidade abdominal	2015
MF615506	Brasil (SC)	Banda coronária	2015
MF615509	Brasil (SC)	Banda coronária	2015
MF615510	Brasil (SC)	Banda coronária	2015
MF615504	Brasil (SC)	Banda coronária	2016
MF615507	Brasil (SC)	Lesão vesicular	2015
MZ456812	Brasil (SC)	Lesão vesicular	2020
MF615503	Brasil (SC)	Pulmão	2016
MF615501	Brasil (SC)	Nasofaringe	2015
MF615502	Brasil (SC)	Nasofaringe e banda coronária	2016
MF615505	Brasil (SC)	Nasofaringe e banda coronária	2016
KY486165	Canadá	Tecidos não especificados	2015
KY486166	Canadá	Tecidos não especificados	2015
KY486156	Canadá	Não informado	2015
KY486157	Canadá	Não informado	2015
KY486158	Canadá	Não informado	2015
KY486159	Canadá	Não informado	2015
KY486160	Canadá	Não informado	2015
KY486161	Canadá	Não informado	2015
KY486162	Canadá	Não informado	2015
KY486163	Canadá	Não informado	2015
KY486164	Canadá	Não informado	2015
KC667560	Canadá	Cérebro	2011
MT360257	Canadá	Não informado	2011
MN233023	Canadá	Não informado	2007
ON843710	Chile	Não informado	2022
KY038016	China	Lesão vesicular	2016
MF460448	China	Lesão vesicular	2016
MF460449	China	Lesão vesicular	2017
MN017170	China	Pulmão	2018
OP562896	China	Tecidos não especificados	2022
MN885796	China	Não informado	2017
MK170056	China	Não informado	2017
OQ162219	China	Não informado	2018
MW713118	China	Não informado	2018
MW713120	China	Não informado	2018
MT840202	China	Não informado	2018
MK252002	China	Não informado	2018
MK170054	China	Não informado	2018

MK170055	China	Não informado	2018
MW713116	China	Não informado	2019
MW713117	China	Não informado	2019
MW713119	China	Não informado	2019
MW713121	China	Não informado	2019
MW713122	China	Não informado	2019
MW713123	China	Não informado	2019
MT457474	China	Não informado	2019
KT321458	China	Não informado	2015
KX173338	China	Não informado	2015
KX173339	China	Não informado	2015
KX751943	China	Não informado	2015
KX173340	China	Não informado	2015
KX751944	China	Não informado	2016
KX751945	China	Não informado	2016
KX751946	China	Não informado	2016
KX377924	China	Não informado	2016
MF893200	China	Não informado	2016
KY747510	China	Não informado	2017
KY747511	China	Não informado	2017
KY747512	China	Não informado	2017
MG765566	China	Não informado	2017
MF189000	China	Não informado	2017
MF189001	China	Não informado	2017
MG765564	China	Não informado	2017
MG765561	China	Não informado	2017
MG765563	China	Não informado	2017
MG765562	China	Não informado	2017
MG765560	China	Não informado	2017
MG765559	China	Não informado	2017
ON024061	China	Não informado	2017
MG765557	China	Não informado	2017
MG428680	China	Não informado	2017
MG428681	China	Não informado	2017
MG428682	China	Não informado	2017
MZ818785	China	Não informado	2017
MG765553	China	Não informado	2017
MG765550	China	Não informado	2017
MG428683	China	Não informado	2017
MH490944	China	Não informado	2017
MG765552	China	Não informado	2017
MG765565	China	Não informado	2017
MG428684	China	Não informado	2017
MG765556	China	Não informado	2017
MG428685	China	Não informado	2017
MG765551	China	Não informado	2017
MF967574	China	Não informado	2017
MG765558	China	Não informado	2017
MH316113	China	Não informado	2017
MH316114	China	Não informado	2017
MH316115	China	Não informado	2017
MH316116	China	Não informado	2017
MH316117	China	Não informado	2017
MG765555	China	Não informado	2017
MG765554	China	Não informado	2017
MN922286	China	Não informado	2017
MH064433	China	Não informado	2017
MH064434	China	Não informado	2017

MH064435	China	Não informado	2017
MH064436	China	Não informado	2017
MN433300	China	Não informado	2018
MH716015	China	Não informado	2018
MK802890	China	Não informado	2018
MK802891	China	Não informado	2018
MK802892	China	Não informado	2018
MK463618	China	Não informado	2018
MK463619	China	Não informado	2018
MH779611	China	Não informado	2018
MK284515	China	Não informado	2018
MH817445	China	Não informado	2018
MH817446	China	Não informado	2018
MH588717	China	Não informado	2018
MN700930	China	Não informado	2018
MK039162	China	Não informado	2018
MN615881	China	Não informado	2018
MK284514	China	Não informado	2018
MK357115	China	Não informado	2018
MK357116	China	Não informado	2018
MK357117	China	Não informado	2018
MN887248	China	Não informado	2019
MN887250	China	Não informado	2019
MT713137	China	Não informado	2019
MN887249	China	Não informado	2019
MN887251	China	Não informado	2019
MW117126	China	Não informado	2020
MW117127	China	Não informado	2020
MW117128	China	Não informado	2020
MW117129	China	Não informado	2020
OM654411	China	Não informado	2021
MN781981	China	Não informado	Não informado
MN781982	China	Não informado	Não informado
MN781983	China	Não informado	Não informado
MN781984	China	Não informado	Não informado
MN423333	China	Não informado	Não informado
MN423334	China	Não informado	Não informado
MN171528	China	Não informado	Não informado
MH844686	China	Não informado	Não informado
MH844687	China	Não informado	Não informado
MH844688	China	Não informado	Não informado
MG983756	China	Não informado	Não informado
MK256736	China	Tonsila, linfonodo, baço e cavidade abdominal	2018
KY419132	China	Não informado	2016
OP714192	China	Não informado	2022
KX857728	Colômbia	Não informado	2016
MN812936	EUA	Não informado	1988
MN812937	EUA	Não informado	1988
MN812938	EUA	Não informado	1989
MN812939	EUA	Não informado	1989
MN812940	EUA	Não informado	1992
MN812941	EUA	Não informado	1997
MN812942	EUA	Não informado	2001
MN812943	EUA	Não informado	2006
MN812944	EUA	Não informado	2007
MN812945	EUA	Não informado	2007
MN812946	EUA	Não informado	2009

MN812947	EUA	Não informado	2013
MN812948	EUA	Não informado	2013
MN812949	EUA	Não informado	2015
MN812952	EUA	Não informado	2015
MN812950	EUA	Não informado	2015
MN812956	EUA	Não informado	2015
MN812951	EUA	Não informado	2015
MN812953	EUA	Não informado	2015
MN812955	EUA	Não informado	2015
MN812957	EUA	Não informado	2015
MN812954	EUA	Não informado	2015
MN812958	EUA	Não informado	2016
MN812959	EUA	Não informado	2017
MN812960	EUA	Não informado	2017
KU359211	EUA	Fezes	2015
KU058182	EUA	Lesão vesicular	2015
KU058183	EUA	Lesão vesicular	2015
MN233030	EUA	Não informado	1990
MN233019	EUA	Não informado	2001
MT360258	EUA	Não informado	2002
MT360259	EUA	Não informado	2012
MT360260	EUA	Não informado	2015
MT360261	EUA	Não informado	2015
MT360262	EUA	Não informado	2015
MN164664	EUA	Não informado	2015
MH634523	EUA	Não informado	2015
MH634524	EUA	Não informado	2015
MH634525	EUA	Não informado	2015
MH634526	EUA	Não informado	2015
MH634527	EUA	Não informado	2015
MH634528	EUA	Não informado	2015
MH634529	EUA	Não informado	2015
MH634530	EUA	Não informado	2015
MH634531	EUA	Não informado	2015
MH634532	EUA	Não informado	2015
MH634533	EUA	Não informado	2015
MH634534	EUA	Não informado	2015
KX019804	EUA	Não informado	2015
KT827250	EUA	Não informado	2015
KT827251	EUA	Não informado	2015
KT827253	EUA	Não informado	2015
MH634506	EUA	Não informado	2017
MH634507	EUA	Não informado	2017
MH634508	EUA	Não informado	2017
MH634509	EUA	Não informado	2017
MH634510	EUA	Não informado	2017
MH634511	EUA	Não informado	2017
MH634512	EUA	Não informado	2017
MH634513	EUA	Não informado	2017
MH634514	EUA	Não informado	2017
MH634515	EUA	Não informado	2017
MH634516	EUA	Não informado	2017
MH634517	EUA	Não informado	2017
MH634518	EUA	Não informado	2017
MH634519	EUA	Não informado	2017
MH634520	EUA	Não informado	2017
MH634521	EUA	Não informado	2017
MH634522	EUA	Não informado	2017

MN233026	EUA	Não informado	1988
MN233027	EUA	Não informado	1988
MN233028	EUA	Não informado	1989
MN233029	EUA	Não informado	1989
MN233031	EUA	Não informado	1992
MN233032	EUA	Não informado	1993
MN233034	EUA	Não informado	1997
MN233033	EUA	Não informado	1999
MN233017	EUA	Não informado	2000
MN233018	EUA	Não informado	2001
MN233020	EUA	Não informado	2005
MN233021	EUA	Não informado	2005
MN233022	EUA	Não informado	2006
KY172968	EUA	Não informado	2014
KU954087	EUA	Não informado	2015
KU051391	EUA	Não informado	2015
KU954088	EUA	Não informado	2015
KU051393	EUA	Não informado	2015
KU051392	EUA	Não informado	2015
KT757280	EUA	Não informado	2015
KX778101	EUA	Não informado	2015
KU051394	EUA	Não informado	2015
KT757281	EUA	Não informado	2015
KU954089	EUA	Não informado	2015
KU954090	EUA	Não informado	2015
KT757282	EUA	Não informado	2015
MK333629	EUA	Não informado	2015
MK333630	EUA	Não informado	2015
MK333631	EUA	Não informado	2015
MN233024	EUA	Não informado	2015
MK333632	EUA	Não informado	2015
MN233025	EUA	Não informado	2015
KX223836	EUA	Não informado	2015
KY618834	EUA	Não informado	2015
KU359210	EUA	Não informado	2015
KU359212	EUA	Não informado	2015
KU359213	EUA	Não informado	2015
KY618835	EUA	Não informado	2016
MK333633	EUA	Não informado	2016
MK333634	EUA	Não informado	2016
MK333635	EUA	Não informado	2016
MK333636	EUA	Não informado	2016
MK333637	EUA	Não informado	2016
KY618836	EUA	Não informado	2016
KY618837	EUA	Não informado	2016
MZ733977	EUA	Não informado	2020
MZ733980	EUA	Não informado	2020
MZ733976	EUA	Não informado	2020
MZ733979	EUA	Não informado	2020
MZ733975	EUA	Não informado	2020
MZ733978	EUA	Não informado	2020
KU954086	EUA	Não informado	Não informado
NC 011349	EUA	Não informado	Não informado
DQ641257	EUA	Não informado	Não informado
KU359214	EUA	Cavidade abdominal	2015
ON369393	México	Não informado	2021
MF416217	Tailândia	Não informado	2016
MF416218	Tailândia	Não informado	2016

MF416219	Tailândia	Não informado	2016
MF416220	Tailândia	Não informado	2016
KY368743	Tailândia	Não informado	2016
KY368744	Tailândia	Não informado	2016
MH704432	Vietnã	Não informado	2018

APÊNDICE II - Dados das 316 sequências de proteína de SVA disponíveis no NCBI *Virus*, com tamanho superior a 2000 aminoácidos, utilizadas no presente trabalho (acesso em 10/01/2024).

Acesso	País de origem	Tipo de amostra	Ano da coleta
AWX33680	Brasil	Banda coronária	2015
AWX33683	Brasil	Banda coronária	2015
AWX33684	Brasil	Banda coronária	2015
AWX33678	Brasil	Banda coronária	2016
AWX33682	Brasil	Cavidade abdominal	2015
AWX33681	Brasil	Lesão vesicular	2015
USZ79719	Brasil	Lesão vesicular	2020
ALD50007	Brasil	Fluido vesicular (informado na publicação)	2015
ALD50008	Brasil	Soro (informado na publicação)	2015
ALD50009	Brasil	Fluido vesicular (informado na publicação)	2015
AWX33675	Brasil	Nasofaringe	2015
AWX33676	Brasil	Nasofaringe e banda coronária	2016
AWX33679	Brasil	Nasofaringe e banda coronária	2016
AWX33677	Brasil	Pulmão	2016
AGM16001	Canadá	Cérebro	2011
QWW20845	Canadá	Não informado	2011
QGL52570	Canadá	Não informado	2007
ARQ19059	Canadá	Não informado	2015
ARQ19060	Canadá	Não informado	2015
ARQ19061	Canadá	Não informado	2015
ARQ19062	Canadá	Não informado	2015
ARQ19063	Canadá	Não informado	2015
ARQ19064	Canadá	Não informado	2015
ARQ19065	Canadá	Não informado	2015
ARQ19066	Canadá	Não informado	2015
ARQ19067	Canadá	Não informado	2015
ARQ19068	Canadá	Tecidos (não especificados)	2015
ARQ19069	Canadá	Tecidos (não especificados)	2015
WBL45177	Chile	Não informado	2022
ARA15253	China	Lesão vesicular	2016
AWU67609	China	Lesão vesicular	2016
AWU67610	China	Lesão vesicular	2017
QOL09545	China	Não informado	2017
QAX58274	China	Não informado	2017
WIM49553	China	Não informado	2018
UTQ10566	China	Não informado	2018
UTQ10568	China	Não informado	2018
QYM90061	China	Não informado	2018
QCC62184	China	Não informado	2018
QAX58272	China	Não informado	2018
QAX58273	China	Não informado	2018

UTQ10564	China	Não informado	2019
UTQ10565	China	Não informado	2019
UTQ10567	China	Não informado	2019
UTQ10569	China	Não informado	2019
UTQ10570	China	Não informado	2019
UTQ10571	China	Não informado	2019
UFQ90508	China	Não informado	2019
QOP59291	China	Não informado	2019
WOZ29597	China	Não informado	2023
ALN69919	China	Não informado	2015
AQM55804	China	Não informado	2015
AQM55805	China	Não informado	2016
AQM55806	China	Não informado	2016
AQM55807	China	Não informado	2016
APA28975	China	Não informado	2016
ATD50366	China	Não informado	2016
ASF79343	China	Não informado	2016
ARR73599	China	Não informado	2017
ARR73600	China	Não informado	2017
ARR73601	China	Não informado	2017
AYA73837	China	Não informado	2017
AVH80610	China	Não informado	2017
AVH80611	China	Não informado	2017
AYA73835	China	Não informado	2017
AYA73832	China	Não informado	2017
AYA73834	China	Não informado	2017
AYA73833	China	Não informado	2017
AYA73831	China	Não informado	2017
AYA73830	China	Não informado	2017
WBR34710	China	Não informado	2017
AYA73828	China	Não informado	2017
AVK39809	China	Não informado	2017
AVK39810	China	Não informado	2017
AVK39811	China	Não informado	2017
UMM61519	China	Não informado	2017
AYA73824	China	Não informado	2017
AYA73821	China	Não informado	2017
AVK39812	China	Não informado	2017
AZB53079	China	Não informado	2017
AYA73823	China	Não informado	2017
AYA73836	China	Não informado	2017
AVK39813	China	Não informado	2017
AYA73827	China	Não informado	2017
AVK39814	China	Não informado	2017
AYA73822	China	Não informado	2017
AWW01209	China	Não informado	2017
AYA73829	China	Não informado	2017
AWU67439	China	Não informado	2017
AWU67440	China	Não informado	2017
AWU67441	China	Não informado	2017
AWU67442	China	Não informado	2017
AWU67443	China	Não informado	2017
AYA73826	China	Não informado	2017
AYA73825	China	Não informado	2017
QOJ79820	China	Não informado	2017
AZV23882	China	Não informado	2017
AZV23883	China	Não informado	2017
AZV23884	China	Não informado	2017

AZV23885	China	Não informado	2017
QGR24460	China	Não informado	2018
QBP34349	China	Não informado	2018
QYY49171	China	Não informado	2018
QMS47479	China	Não informado	2018
QYY49165	China	Não informado	2018
QMS47473	China	Não informado	2018
QEP29104	China	Não informado	2018
QEP29105	China	Não informado	2018
QEP29106	China	Não informado	2018
QBI89945	China	Não informado	2018
QBI89946	China	Não informado	2018
QBS54794	China	Não informado	2018
QMS47480	China	Não informado	2018
QDK54852	China	Não informado	2018
QBZ96258	China	Não informado	2018
QBZ96259	China	Não informado	2018
QBK95561	China	Não informado	2018
QYY49172	China	Não informado	2018
QNC49834	China	Não informado	2018
QYY49174	China	Não informado	2018
QYY49173	China	Não informado	2018
QMS47481	China	Não informado	2018
QCP57166	China	Não informado	2018
UVG61495	China	Não informado	2018
QMS47482	China	Não informado	2018
QMS47475	China	Não informado	2018
QYY49167	China	Não informado	2018
QKV49433	China	Não informado	2018
QMS47476	China	Não informado	2018
QYY49168	China	Não informado	2018
QDK54851	China	Não informado	2018
QEA69442	China	Não informado	2018
QEA69443	China	Não informado	2018
QEA69444	China	Não informado	2018
QYY49166	China	Não informado	2018
QMS47474	China	Não informado	2018
QYY49175	China	Não informado	2019
QYY49170	China	Não informado	2019
QMS47478	China	Não informado	2019
QMS47477	China	Não informado	2019
QMS47483	China	Não informado	2019
QYY49169	China	Não informado	2019
QKV28187	China	Não informado	2019
QMS47484	China	Não informado	2019
QYY49176	China	Não informado	2019
QKV28189	China	Não informado	2019
QZL11397	China	Não informado	2019
QKV28188	China	Não informado	2019
QKV28190	China	Não informado	2019
QXN58110	China	Não informado	2020
QXN58111	China	Não informado	2020
QXN58112	China	Não informado	2020
QXN58113	China	Não informado	2020
UVG61490	China	Não informado	2020
UVG61491	China	Não informado	2020
UVG61496	China	Não informado	2020
UVG61494	China	Não informado	2020

UVG61492	China	Não informado	2020
UVG61493	China	Não informado	2020
UVG61497	China	Não informado	2020
WEL12798	China	Não informado	2020
UMF44225	China	Não informado	2021
UVG61498	China	Não informado	2021
UVG61499	China	Não informado	2021
WGT80121	China	Não informado	2022
QNT22068	China	Não informado	Não informado
QNT22069	China	Não informado	Não informado
QNT22070	China	Não informado	Não informado
QNT22071	China	Não informado	Não informado
QKE43320	China	Não informado	Não informado
QKE43321	China	Não informado	Não informado
QAY30712	China	Não informado	Não informado
QAY30714	China	Não informado	Não informado
QJF53926	China	Pulmão	2018
WOV89872	China	Tecidos (não especificados)	2020
WES13129	China	Tecidos (não especificados)	2022
AZL87715	China	Tonsila, linfonodo, baço e cavidade abdominal	2018
APZ85840	Colômbia	Não informado	2016
AMV75278	EUA	Cavidade abdominal	2015
AMV75275	EUA	Fezes	2015
ANK58781	EUA	Lesão vesicular	2015
ANK58782	EUA	Lesão vesicular	2015
QGL52577	EUA	Não informado	1990
QGL52566	EUA	Não informado	2001
QWW20846	EUA	Não informado	2002
QWW20847	EUA	Não informado	2012
ANN83864	EUA	Não informado	2015
ANN83867	EUA	Não informado	2015
ANN83868	EUA	Não informado	2015
ANN83865	EUA	Não informado	2015
QWW20848	EUA	Não informado	2015
QWW20849	EUA	Não informado	2015
QWW20850	EUA	Não informado	2015
QGX86393	EUA	Não informado	2015
QDZ58648	EUA	Não informado	2015
QDZ58649	EUA	Não informado	2015
QDZ58650	EUA	Não informado	2015
QDZ58651	EUA	Não informado	2015
QDZ58652	EUA	Não informado	2015
QDZ58653	EUA	Não informado	2015
QDZ58654	EUA	Não informado	2015
QDZ58655	EUA	Não informado	2015
QDZ58656	EUA	Não informado	2015
QDZ58657	EUA	Não informado	2015
QDZ58658	EUA	Não informado	2015
QDZ58659	EUA	Não informado	2015
ANX99825	EUA	Não informado	2015
ANN83866	EUA	Não informado	2015
QDZ58631	EUA	Não informado	2017
QDZ58632	EUA	Não informado	2017
QDZ58633	EUA	Não informado	2017
QDZ58634	EUA	Não informado	2017
QDZ58635	EUA	Não informado	2017
QDZ58636	EUA	Não informado	2017

QDZ58637	EUA	Não informado	2017
QDZ58638	EUA	Não informado	2017
QDZ58639	EUA	Não informado	2017
QDZ58640	EUA	Não informado	2017
QDZ58641	EUA	Não informado	2017
QDZ58642	EUA	Não informado	2017
QDZ58643	EUA	Não informado	2017
QDZ58644	EUA	Não informado	2017
QDZ58645	EUA	Não informado	2017
QDZ58646	EUA	Não informado	2017
QDZ58647	EUA	Não informado	2017
QGL52573	EUA	Não informado	1988
QHF16679	EUA	Não informado	1988
QHF16680	EUA	Não informado	1988
QGL52574	EUA	Não informado	1988
QGL52575	EUA	Não informado	1989
QHF16681	EUA	Não informado	1989
QGL52576	EUA	Não informado	1989
QHF16682	EUA	Não informado	1989
QHF16683	EUA	Não informado	1992
QGL52578	EUA	Não informado	1992
QGL52579	EUA	Não informado	1993
QHF16684	EUA	Não informado	1997
QGL52581	EUA	Não informado	1997
QGL52580	EUA	Não informado	1999
QGL52564	EUA	Não informado	2000
QHF16685	EUA	Não informado	2001
QGL52565	EUA	Não informado	2001
QGL52567	EUA	Não informado	2005
QGL52568	EUA	Não informado	2005
QHF16686	EUA	Não informado	2006
QGL52569	EUA	Não informado	2006
QHF16687	EUA	Não informado	2007
QHF16688	EUA	Não informado	2007
QHF16689	EUA	Não informado	2009
UXT88225	EUA	Não informado	2013
QHF16690	EUA	Não informado	2013
QHF16691	EUA	Não informado	2013
AQA27243	EUA	Não informado	2014
QHF16692	EUA	Não informado	2015
QHF16695	EUA	Não informado	2015
QHF16693	EUA	Não informado	2015
AQM55550	EUA	Não informado	2015
ANN87789	EUA	Não informado	2015
AQM55551	EUA	Não informado	2015
ANN87791	EUA	Não informado	2015
ANN87790	EUA	Não informado	2015
ALF40133	EUA	Não informado	2015
APY18927	EUA	Não informado	2015
ANN87792	EUA	Não informado	2015
ALF40134	EUA	Não informado	2015
AQM55552	EUA	Não informado	2015
AQM55553	EUA	Não informado	2015
ALF40135	EUA	Não informado	2015
QGL52582	EUA	Não informado	2015
QGL52583	EUA	Não informado	2015
QGL52584	EUA	Não informado	2015
QHF16699	EUA	Não informado	2015

QGL52571	EUA	Não informado	2015
QHF16694	EUA	Não informado	2015
QHF16696	EUA	Não informado	2015
QGL52585	EUA	Não informado	2015
QGL52572	EUA	Não informado	2015
APW78915	EUA	Não informado	2015
AUW36853	EUA	Não informado	2015
AMV75274	EUA	Não informado	2015
AMV75276	EUA	Não informado	2015
AMV75277	EUA	Não informado	2015
QHF16698	EUA	Não informado	2015
QHF16700	EUA	Não informado	2015
QHF16697	EUA	Não informado	2015
AUW36854	EUA	Não informado	2016
QGL52586	EUA	Não informado	2016
QGL52587	EUA	Não informado	2016
QGL52588	EUA	Não informado	2016
QGL52589	EUA	Não informado	2016
QGL52590	EUA	Não informado	2016
AUW36855	EUA	Não informado	2016
AUW36856	EUA	Não informado	2016
QHF16701	EUA	Não informado	2016
QHF16702	EUA	Não informado	2017
QHF16703	EUA	Não informado	2017
UEW62690	EUA	Não informado	2020
UEW62693	EUA	Não informado	2020
UEW62689	EUA	Não informado	2020
UEW62692	EUA	Não informado	2020
UEW62688	EUA	Não informado	2020
UEW62691	EUA	Não informado	2020
YP_002268402	EUA	Não informado	Não informado
AQM55549	EUA	Não informado	Não informado
ABG23522	EUA	Não informado	Não informado
UXX39431	México	Não informado	2021
QID77682	Não informado	Não informado	Não informado
Q155Z9	Não informado	Não informado	Não informado
AYQ58338	Não informado	Não informado	Não informado
AVG70999	Tailândia	Não informado	2016
AVG71000	Tailândia	Não informado	2016
AVG71001	Tailândia	Não informado	2016
AVG71002	Tailândia	Não informado	2016
ARQ31641	Tailândia	Não informado	2016
ARQ31642	Tailândia	Não informado	2016
AXS67918	Vietnã	Não informado	2018

APÊNDICE III - Dados das sequências brasileiras de SVA disponíveis no NCBI *Virus*, com tamanho superior a 7000 pares de base e 2000 aminoácidos, utilizadas no presente trabalho (acesso em 30/10/2023 e 10/01/2024, respectivamente).

Sequência de nucleotídeos	Sequência de proteínas	País de origem	Tipo de amostra	Ano da coleta
KR063107	ALD50007.1	Brasil (MG)	Fluido vesicular (informado na publicação)	2015
KR063108	ALD50008.1	Brasil (MG)	Soro (informado na publicação)	2015
KR063109	ALD50009.1	Brasil (GO)	Fluido vesicular (informado na publicação)	2015
MF615501	AWX33675.1	Brasil (SC)	Nasofaringe	2015
MF615502	AWX33676.1	Brasil (SC)	Nasofaringe e banda coronária	2016
MF615503	AWX33677.1	Brasil (SC)	Pulmão	2016
MF615504	AWX33678.1	Brasil (SC)	Banda coronária	2016
MF615505	AWX33679.1	Brasil (SC)	Nasofaringe e banda coronária	2016
MF615506	AWX33680.1	Brasil (SC)	Banda coronária	2015
MF615507	AWX33681.1	Brasil (SC)	Lesão vesicular	2015
MF615508	AWX33682.1	Brasil (SC)	Cavidade abdominal	2015
MF615509	AWX33683.1	Brasil (SC)	Banda coronária	2015
MF615510	AWX33684.1	Brasil (SC)	Banda coronária	2015
MZ456812	USZ79719.1	Brasil (SC)	Lesão vesicular	2020