



**UNIVERSIDADE DO ESTADO DA BAHIA**  
**DEPARTAMENTO DE CIÊNCIAS EXATAS E DA TERRA**  
**CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO**

**MAURICIO SOUZA MENEZES**

**FERRAMENTA COMPUTACIONAL PARA ESTUDO DA EVOLUÇÃO DE ESPÉCIES**  
**VIRAIS BASEADO NO USO DE CÓDONS**

**SALVADOR, BAHIA, BRASIL**

**2023**

MAURICIO SOUZA MENEZES

FERRAMENTA COMPUTACIONAL PARA ESTUDO DA EVOLUÇÃO DE ESPÉCIES  
VIRAIS BASEADO NO USO DE CÓDONS

Monografia apresentada ao curso de Sistemas de Informação do Departamento de Ciências Exatas e da Terra da Universidade do Estado da Bahia - UNEB, como requisito parcial à obtenção do grau de bacharel em Sistemas de Informação. Área de Concentração: Ciência da Computação

Orientador: PhD Diego Gervasio Frias Suárez

Co-Orientador: PhD Vagner Fonseca

SALVADOR, BAHIA, BRASIL

2023

## **Termo de Anuência do Orientador**

Declaro para os devidos fins que li e revisei este trabalho e atesto sua qualidade como resultado parcial desta monografia. Confirmo que o referencial teórico apresentado é completo e suficiente para fundamentar os objetivos propostos e que o início da modelagem ou desenvolvimento da solução proposta está apresentado neste texto com plano de validação completo, incluindo métricas e indicadores, revisado e atualizado. Atesto que este trabalho está apto a ser apresentado na pré-banca final da disciplina Trabalho de Conclusão de Curso I.

---

PhD Diego Gervasio Frias Suárez

Dedico este trabalho, com muito amor, a minha  
rainha, Miriam Souza Menezes

## **AGRADECIMENTOS**

Agradeço a Deus pela vida e por me guiar nos caminhos certos; Agradeço aos meus pais, Mauricio Porto e Miriam Souza, pela criação e por todo o apoio que me deram; Agradeço também ao meu irmão (mesmo sem merecer. . .); Agradeço a minha namorada, Yasmim Arrais, por todo o apoio, conversas e momentos em que me tranquilizou; Agradeço ao meu orientador, Diego Frias, pela amizade, paciência e atenção dada. Agradeço a todos os colegas de curso, em especial aos amigos Joílson Argolo e Marcelo Henrique, que estiveram sempre próximos durante toda essa caminhada. Agradeço ao meu amigo, Alexandre Aquiles, por me ensinar ainda mais, que ajudar ao próximo é essencial em todos os momentos da nossa vida.

## RESUMO

Este trabalho tem como objetivo principal o desenvolvimento de um modelo para a análise de genomas virais, baseado no uso de códons. Essa ferramenta se propõe a ser uma importante ferramenta para a análise da evolução de espécies, utilizando sequências genômicas do SARS-COV-2 como base de estudo. A implementação desse modelo visa proporcionar maior eficiência computacional e alcançar resultados mais precisos. Adicionalmente, a ferramenta será capaz de apresentar visualizações gráficas dos resultados obtidos, facilitando a interpretação dos dados e auxiliando na tomada de decisões científicas. Espera-se que essa abordagem proporcione insights valiosos sobre a evolução de espécies virais, contribuindo para o avanço da virologia e da genômica comparativa. Os resultados obtidos serão analisados com o objetivo de demonstrar a eficácia dessa ferramenta na compreensão dos padrões evolutivos em espécies virais, tornando-a uma promissora aliada para pesquisadores e profissionais da área.

**Palavras-chave:** Bioinformática. Códons. Filogenia. Viral.

## **ABSTRACT**

This work aims to develop a model for the analysis of viral genomes based on the use of codons, which will serve as a tool for studying the evolution of species using genomic sequences of SARS-CoV-2. It is expected that this approach will enable a more efficient computational process and yield improved results. Additionally, the tool will provide graphical visualization of the results, facilitating data interpretation and supporting scientific decision-making. Through the application of this tool, valuable insights into the evolution of viral species are anticipated, contributing to advancements in virology and comparative genomics. The obtained results are expected to demonstrate the effectiveness of the tool in analyzing and understanding evolutionary patterns in viral species, making it a promising resource for researchers and professionals in the field.

**Keywords:** Bioinformatics. Codons. Phylogeny. Viral.

## LISTA DE FIGURAS

Figura 1 – Estrutura do DNA. . . . .	16
Figura 2 – Estrutura do coronavírus. . . . .	17



## **LISTA DE TABELAS**

## **LISTA DE ABREVIATURAS E SIGLAS**

BV-BRC	Bacterial and Viral Bioinformatics Resource Center
COVID-19	Coronavirus Disease 2019
DNA	Ácido Desoxirribonucleico
DSR	Design Science Research
ML	Máxima Verossimilhança
mRNA	Ácido Ribonucleico Mensageiro
RNA	Ácido Ribonucleico
tRNA	Ácido Ribonucleico Transportador

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>11</b>
<b>2</b>	<b>CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA . . . . .</b>	<b>15</b>
<b>2.1</b>	<b>Biologia Molecular . . . . .</b>	<b>15</b>
<b>2.2</b>	<b>Virus . . . . .</b>	<b>16</b>
2.2.1	SARS-CoV-2 . . . . .	17
<b>2.3</b>	<b>Filogenia . . . . .</b>	<b>17</b>
<b>3</b>	<b>CAPÍTULO 3 - DESCRIÇÃO DO PROJETO . . . . .</b>	<b>19</b>
<b>3.1</b>	<b>Plano de Implementação . . . . .</b>	<b>19</b>
<b>4</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>21</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>22</b>

## 1 INTRODUÇÃO

Os desafios impostos pela pandemia do COVID-19 incluíram a falta de conhecimento suficiente para a compreensão da importância das ameaças biológicas e para a preparação médica, apesar dos avanços científicos e tecnológicos. O conhecimento prévio sobre os agentes biológicos com potencial para causar pandemias pode melhorar substancialmente nossa preparação pré-pandemia (1).

Diante disso, a bioinformática, que é a junção de métodos computacionais e técnicas estatísticas com o objetivo de extrair informações de dados biológicos brutos, desempenha um papel fundamental na interpretação de dados genômicos e na compreensão de processos evolutivos. Segundo Hall e Barlow(2) os métodos filogenéticos podem ser usados para analisar os dados da sequência de nucleotídeos de forma que a ordem de descendência de cepas relacionadas possa ser determinada. Quando associada à análise filogenética apropriada, a epidemiologia molecular tem o potencial de elucidar os mecanismos que levam a surtos microbianos e epidemias.”

A reconstrução filogenética é uma das abordagens amplamente utilizadas na análise da evolução de espécies, que permite investigar as relações evolutivas entre diferentes linhagens de vírus. Essas observações são realizadas com base em dados como sequências de Ácido Desoxirribonucleico (DNA) e Ácido Ribonucleico (RNA). Essas sequências são formadas por blocos fundamentais chamados de nucleotídeos, que são compostos por uma base nitrogenada, um açúcar e um grupo fosfato. As bases presentes nos nucleotídeos do DNA são adenina (A), timina (T), citosina (C) e guanina (G), enquanto no RNA a base timina é substituída pela uracila (U) (??).

Uma das principais formas de análise filogenética é realizada através da árvore filogenética, onde são representadas as relações evolutivas entre um conjunto de espécies. De acordo com Morrison(3) elas tem função importante porque apresentam de forma sucinta e particular a evolução dos descendentes partindo de ancestrais em comum.

A semelhança genética entre vários vírus infecciosos e mortais fornece uma visão do fato de que o RNA é a chave para discernir e marcar os possíveis patógenos que podem causar uma pandemia. Embora um padrão geral e motivos conservados possam ser observados em

ancestrais imediatos, as regiões não conservadas das sequências são o resultado da acumulação de mutações, seja por inserção ou deleção de um ou vários nucleotídeos ou por substituição pontual de um nucleotídeo por outro. A fonte principal de mutações em vírus são percalços na replicação e a recombinação de RNA (1).

Apesar da utilidade da filogenética e dos softwares comerciais e públicos disponíveis para análises filogenéticas, os métodos filogenéticos são muitas vezes aplicados de forma inadequada. Mesmo quando aplicados adequadamente, são mal explicados e, portanto, mal compreendidos. (2, p. 1) Além disso, por trabalhar com grandes quantidades de dados, os métodos utilizados devem ser avaliados também em relação ao seu custo computacional.

Na busca de trabalhos relacionados, vários métodos foram encontrados, e a seguir são apresentados.

O método de Máxima Verossimilhança (ML) (ou *Maximum Likelihood*), não é exclusivo da filogenia, mas sim uma abordagem estatística. A sua aplicação em filogenia consiste em avaliar a probabilidade de que o modelo de evolução escolhido gere os dados observados, que são por exemplo, características de um organismo. Essa proposta foi utilizada por (4, 1, 5, 6, 7, 8, 9, 10).

Já em (11, 12), foi usada a inferência bayesiana, que é fundamentada no teorema de Bayes, que permite a atualização das probabilidades a priori para probabilidades a posteriori à medida que novas evidências são incorporadas.

Além desses, (13) utilizou a junção de vizinhos (ou *Neighbor-Joining*), que é baseado em uma abordagem heurística que visa construir uma árvore filogenética a partir de uma matriz de distância entre as sequências estudadas. O trabalho de (14) expôs o Frequency Chaos Game Representation e (15) a floresta aleatória. Por fim, (16) comparou três modelos para reconstrução de árvores filogenéticas: junção de vizinhos; ML e inferência bayesiana.

As soluções até então desenvolvidas, são guiadas pela reconstrução das árvores filogenéticas construídas a partir das mutações de nucleotídeos. Neste aspecto, as ferramentas disponíveis não oferecem uma aplicação no contexto de árvores reconstruídas com distâncias obtidas a partir da diferença do uso de códons, que são sequências de três nucleotídeos responsáveis pela codificação dos aminoácidos nas proteínas. Os códons desempenham um papel crucial na determinação da função e estrutura das proteínas, e alterações nos códons podem

resultar em mudanças significativas nas características fenotípicas dos vírus. Necessita-se então, de pesquisas e desenvolvimento de ferramentas que realizem uma classificação de sequências genéticas com base no uso/frequência de códons.

Com base no problema de pesquisa proposto, foram construídos os objetivos que deveriam ser atingidos, os mesmos são apresentados a seguir:

- Objetivo Geral
  - (i) Desenvolver e validar um novo método de análise da evolução molecular viral.
- Objetivos Específicos
  - (i) Coletar os dados necessários para aplicar na ratificação do método.
  - (ii) Definir um modelo para validação do método proposto.
  - (iii) Desenvolver uma ferramenta para caracterizar/validar o método.
  - (iv) Disponibilizar o modelo como uma ferramenta web de fácil acesso.
  - (v) Realizar a comparação da performance computacional no novo método com algum dos métodos do estado da arte.

Atingir o objetivo de desenvolver um método de construção de árvores com base nas distâncias obtidas a partir da diferença do uso de códons, contribuiria com a tarefa de classificação de cepas para a vigilância sanitária, especialmente na descoberta de novas cepas emergentes com potenciais pandêmicos. Ademais, é também importante dispor de alternativas à filogenia molecular atualmente utilizada, para gerar informações de outro ponto de vista e-ou para servir de referência aos métodos filogenéticos. Os métodos atuais ainda demandam de um alto custo computacional, sendo assim, existe a necessidade de desenvolver outros mais baratos e que possam suportar o volume crescente de dados (sequências). Posto isso, o projeto visa apresentar um método que seja capaz de realizar classificações, com um custo computacional baixo, em relação a outros métodos, e que possa apresentar, do ponto de vista científico, alternativas de comparação com outras técnicas já existentes.

Um ponto importante para a obtenção dos objetivos deste trabalho está relacionada a definição da metodologia que servirá como alicerce. Com a proposta de desenvolver e validar um método de análise da evolução molecular de vírus com base no uso de códons, a metodologia escolhida para isso é o Design Science Research (DSR). Essa metodologia, proporciona um framework teórico e prático para a criação de artefatos inovadores, como métodos, modelos ou frameworks, visando resolver problemas específicos.(17)

Para a obtenção de sucesso ao utilizar o DSR os seguintes passos serão seguidos:

1. Identificação do problema e definição dos objetivos.
2. Desenvolvimento do artefatos.
3. Avaliação do artefato.
4. Apresentar contribuições científicas.

Também será utilizada análises quantitativas, ou seja, medidas estatísticas para mensurar e comparar os resultados obtidos.

A pesquisa quantitativa só tem sentido quando há um problema muito bem definido e há informação e teoria a respeito do objeto de conhecimento, entendido aqui como o foco da pesquisa e/ou aquilo que se quer estudar. Esclarecendo mais, só se faz pesquisa de natureza quantitativa quando se conhece as qualidades e se tem controle do que se vai pesquisar.(18)

Os pontos a seguir serão realizados durante o desenvolvimento do projeto:

- Coleta de sequências que serão utilizadas.
- Tratamento necessário dos dados.
- Avaliação de desempenho do modelo.
- Análise comparativa com modelos convencionais.
- Disponibilização do modelo como ferramenta web.

## **2 CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA**

### **2.1 BIOLOGIA MOLECULAR**

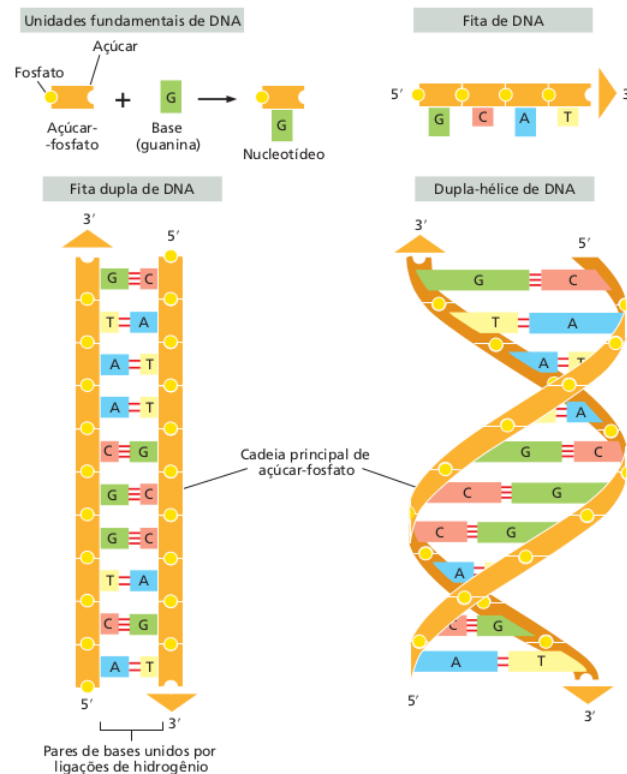
A Biologia Molecular é um ramo da biologia que lida e investiga os processos e mecanismos moleculares relacionados à estrutura, função e interações das biomoléculas presentes nos organismos vivos. Consiste em principalmente em estudar as interações entre os vários sistemas da célula, partindo da relação entre o DNA, RNA e a síntese de proteínas, e o modo como essas interações são reguladas.

É importante entender a estrutura do DNA vista na Figura 1, está é uma molécula em forma de dupla hélice que carrega a informação genética em organismos vivos. Ela é composta por duas cadeias polinucleotídicas complementares enroladas em torno de um eixo central. Cada cadeia é composta por uma sequência de nucleotídeos, que consistem em uma pentose (a desoxirribose), um grupo fosfato e uma base nitrogenada (adenina, timina, citosina ou guanina). A estrutura do DNA é mantida por pontes de hidrogênio entre as bases complementares, com a adenina pareando sempre com a timina e a citosina pareando sempre com a guanina.

As informações contidas no DNA é copiada em uma molécula de RNA, esse processo é conhecido como transcrição. A transcrição ocorre no núcleo das células e envolve a separação das duas fitas do DNA e o pareamento de nucleotídeos complementares para sintetizar uma molécula de Ácido Ribonucleico Mensageiro (mRNA). O mRNA é uma cópia do DNA que carrega a sequência de bases nitrogenadas correspondente a um gene específico. Após isso, ocorre o processo de tradução onde a sequência de bases nitrogenadas do mRNA é utilizada para sintetizar proteínas. A tradução ocorre nos ribossomos, presentes no citoplasma celular. Durante a tradução, o mRNA é lido em grupos de três bases, chamados de códon. Cada códon especifica um aminoácido distinto. Os aminoácidos são transportados para o ribossomo por moléculas de Ácido Ribonucleico Transportador (tRNA), que possuem um anticódon complementar ao códon do mRNA. À medida que o ribossomo percorre o mRNA, os aminoácidos são ligados em uma sequência específica, formando uma cadeia polipeptídica que será dobrada e modificada para se



Figura 1 – Estrutura do DNA.



Fonte: Retirada de Alberts et al.(19)

tornar uma proteína funcional (19).

## 2.2 VIRUS

Os vírus são agentes infecciosos compostos por uma cápsula proteica (capsídeo) que envolve seu material genético, que pode ser DNA ou RNA. A estrutura viral varia entre diferentes tipos de vírus, mas de modo geral, eles consistem em uma cápsula proteica chamada capsídeo, que envolve o material genético viral. O capsídeo pode apresentar diferentes formas, como hélices, icosaedros ou formas complexas. Além do capsídeo, alguns vírus possuem uma camada lipídica chamada envelope viral, que é derivada da membrana da célula hospedeira e contém glicoproteínas virais que são importantes para a entrada do vírus nas células hospedeiras (20). O ciclo e vida viral é conjunto de etapas que um vírus passa para se reproduzir e infectar novas células. Esse ciclo pode variar entre diferentes tipos de vírus, mas geralmente envolve as seguintes etapas:

1. Adsorção: o vírus se liga especificamente a receptores na superfície da célula hospedeira.
2. Penetração: o vírus entra na célula hospedeira, liberando seu material genético.
3. Replicação e síntese de proteínas virais: o material genético viral é replicado e transcritas

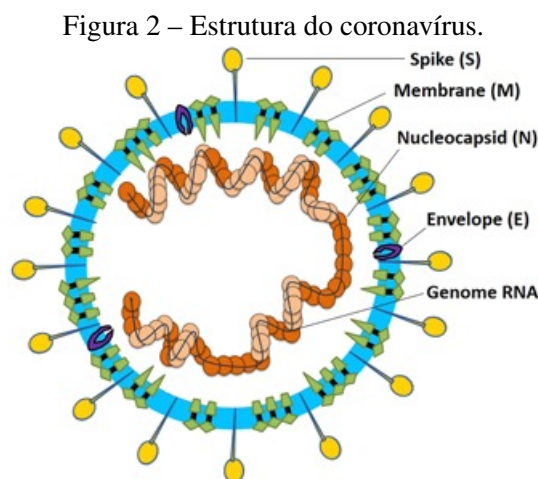
em moléculas de mRNA, que são utilizadas para a síntese de proteínas virais

4. Montagem: as proteínas virais se unem para formar novas partículas virais
5. Liberação: as novas partículas virais são liberadas da célula hospedeira, podendo ocorrer por lise celular ou por brotamento

### 2.2.1 SARS-CoV-2

O SARS-CoV-2 é um vírus da família Coronaviridae, que causa a doença chamada Coronavirus Disease 2019 (COVID-19). Ele foi identificado pela primeira vez em dezembro de 2019 na cidade de Wuhan, na província de Hubei, na China, e desde então se espalhou para todo o mundo, resultando em uma pandemia global (21, 22).

O SARS-CoV-2 possui uma estrutura viral apresentada na Figura 2, característica dos coronavírus. Ele é composto por uma partícula viral esférica, com um envelope lipídico que envolve seu material genético. A estrutura do vírus inclui proteínas de espículas na sua superfície, conhecidas como proteína spike (S), que são responsáveis pela ligação do vírus às células hospedeiras. Além disso, o SARS-CoV-2 possui proteínas de membrana (M), envelope (E) e nucleocapsídeo (N), que desempenham papéis importantes na estrutura e na replicação viral.



Fonte: Retirada de Li et al.(23)

## 2.3 FILOGENIA

A filogenia é uma disciplina da biologia que estuda as relações evolutivas entre organismos, buscando reconstruir a história evolutiva e a ancestralidade comum. A filogené-

tica molecular é uma abordagem utilizada para inferir a filogenia com base em informações moleculares, como sequências de DNA, RNA e proteínas(24).

A construção de árvores filogenéticas é um aspecto fundamental da filogenética molecular. Existem vários métodos utilizados para construir árvores filogenéticas, que podem ser classificados em dois grupos principais: métodos baseados em distância e métodos baseados em caracteres. Os métodos baseados em distância medem a similaridade ou a dissimilaridade entre sequências moleculares e constroem árvores filogenéticas com base nessas medidas. Alguns exemplos de métodos baseados em distância incluem o método de Neighbor Joining (NJ) e o método de Mínima Evolução (ME). Por outro lado, os métodos baseados em caracteres analisam as mudanças nos caracteres moleculares ao longo do tempo para inferir as relações filogenéticas. Exemplos de métodos baseados em caracteres são o método de Máxima Parcimônia (MP) e o método de Inferência Bayesiana(25).

### 3 CAPÍTULO 3 - DESCRIÇÃO DO PROJETO

A seguir, serão apresentadas a metodologia adotada neste estudo, bem como as etapas detalhadas de sua implementação.

#### 3.1 PLANO DE IMPLEMENTAÇÃO

- **Desenvolvimento do Modelo:** Para a realização do desenvolvimento do modelo serão realizados os seguintes passos:
  - Definição do problema a ser solucionado.
  - Levantamento dos requisitos.
  - Coleta dos dados relevantes para a implementação.
  - Análise exploratória dos dados.
  - Pré-processamento dos dados obtidos.
  - Seleção e engenharia de recursos.
  - Escolha do algoritmo.
  - Treinamento do modelo.
  - Validação e avaliação.
  - Ajustes e otimização.
  - Teste do modelo.
- **Montagem do Dataset:** O download das sequências necessárias para a realização do projeto será realizada através do site Bacterial and Viral Bioinformatics Resource Center (BV-BRC). O BV-BRC, é um centro de recursos de bioinformática dedicado ao estudo e análise de bactérias e vírus. O site também disponibiliza uma coleção abrangente de banco de dados, incluindo sequências genômicas, anotações funcionais, informações de expressão gênica e estruturas tridimensionais. O acesso aos bancos de dados é dado por meio de uma interface amigável, onde é possível realizar pesquisas avançadas. Também será realizado os tratamentos necessários para que os dados obtidos estejam padronizados no formato esperado pelo modelo a ser desenvolvido.
- **Análise comparativa entre o método proposto e outro método existente:** Será realizada

uma análise com um conjunto de dados, onde será realizado análises estatísticas para verificação de melhorias, ou não, do novo método proposto.

#### **4 CONSIDERAÇÕES FINAIS**

## REFERÊNCIAS

- 1 BEHL, A.; NAIR, A.; MOHAGAONKAR, S.; YADAV, P.; GAMBHIR, K.; TYAGI, N.; SHARMA, R. K.; BUTOLA, B. S.; SHARMA, N. Threat, challenges, and preparedness for future pandemics: A descriptive review of phylogenetic analysis based predictions. **Infection, Genetics and Evolution**, v. 98, p. 105217, mar. 2022. ISSN 15671348. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1567134822000144>>.
- 2 HALL, B. G.; BARLOW, M. Phylogenetic analysis as a tool in molecular epidemiology of infectious diseases. **Annals of Epidemiology**, v. 16, n. 3, p. 157–169, 2006. ISSN 1047-2797. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1047279705001080>>.
- 3 MORRISON, D. A. Tree Thinking: An Introduction to Phylogenetic Biology. David A. Baum and Stacey D. Smith. **Systematic Biology**, v. 62, n. 4, p. 634–637, 05 2013. ISSN 1063-5157. Disponível em: <<https://doi.org/10.1093/sysbio/syt026>>.
- 4 FALL, A.; ELAWAR, F.; HODCROFT, E. B.; JALLOW, M. M.; TOURE, C. T.; BARRY, M. A.; KIORI, D. E.; SY, S.; DIAW, Y.; GOUDIABY, D.; NIANG, M. N.; DIA, N. Genetic diversity and evolutionary dynamics of respiratory syncytial virus over eleven consecutive years of surveillance in Senegal. **Infection, Genetics and Evolution**, v. 91, p. 104864, jul. 2021. ISSN 15671348. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1567134821001611>>.
- 5 SHABBIR, M. Z.; RAHMAN, A.-u.; MUNIR, M. A comprehensive global perspective on phylogenomics and evolutionary dynamics of Small ruminant morbillivirus. **Scientific Reports**, v. 10, n. 1, p. 17, dez. 2020. ISSN 2045-2322. Disponível em: <<http://www.nature.com/articles/s41598-019-54714-w>>.
- 6 HUDU, S. A.; NIAZLIN, M. T.; NORDIN, S. A.; HARMAL, N. S.; TAN, S. S.; OMAR, H.; SHAHAR, H.; MUTALIB, N. A.; SEKAWI, Z. Hepatitis E virus isolated from chronic hepatitis B patients in Malaysia: Sequences analysis and genetic diversity suggest zoonotic origin. **Alexandria Journal of Medicine**, v. 54, n. 4, p. 487–494, dez. 2018. ISSN 2090-5068, 2090-5076. Disponível em: <<https://www.tandfonline.com/doi/full/10.1016/j.ajme.2017.07.003>>.
- 7 SALLARD, E.; HALLOY, J.; CASANE, D.; DECROLY, E.; HELDEN, J. van. Tracing the origins of SARS-COV-2 in coronavirus phylogenies: a review. **Environmental Chemistry Letters**, v. 19, n. 2, p. 769–785, abr. 2021. ISSN 1610-3653, 1610-3661. Disponível em: <<https://link.springer.com/10.1007/s10311-020-01151-1>>.
- 8 PAEZ-ESPINO, D.; ZHOU, J.; ROUX, S.; NAYFACH, S.; PAVLOPOULOS, G. A.; SCHULZ, F.; MCMAHON, K. D.; WALSH, D.; WOYKE, T.; IVANOVA, N. N.; ELOE-FADROSH, E. A.; TRINGE, S. G.; KYRPIDES, N. C. Diversity, evolution, and classification of virophages uncovered through global metagenomics. **Microbiome**, v. 7, n. 1, p. 157, dez. 2019. ISSN 2049-2618. Disponível em: <<https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-019-0768-5>>.
- 9 TANG, X.; YING, R.; YAO, X.; LI, G.; WU, C.; TANG, Y.; LI, Z.; KUANG, B.; WU, F.; CHI, C.; DU, X.; QIN, Y.; GAO, S.; HU, S.; MA, J.; LIU, T.; PANG, X.; WANG, J.; ZHAO, G.; TAN, W.; ZHANG, Y.; LU, X.; LU, J. Evolutionary analysis and lineage designation of

SARS-CoV-2 genomes. **Science Bulletin**, v. 66, n. 22, p. 2297–2311, nov. 2021. ISSN 20959273. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2095927321001250>>.

10 CHO, M.; MIN, X.; SON, H. S. Analysis of evolutionary and genetic patterns in structural genes of primate lentiviruses. **Genes & Genomics**, v. 44, n. 7, p. 773–791, jul. 2022. ISSN 1976-9571, 2092-9293. Disponível em: <<https://link.springer.com/10.1007/s13258-022-01257-6>>.

11 YIN, Y.; HE, K.; WU, B.; XU, M.; DU, L.; LIU, W.; LIAO, P.; LIU, Y.; HE, M. A systematic genotype and subgenotype re-ranking of hepatitis B virus under a novel classification standard. **Heliyon**, v. 5, n. 10, p. e02556, out. 2019. ISSN 24058440. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405844019362164>>.

12 BEDOYA-PILOZO, C. H.; MAGÜES, L. G. M.; ESPINOSA-GARCÍA, M.; SÁNCHEZ, M.; VALDIVIEZO, J. V. P.; MOLINA, D.; IBARRA, M. A.; QUIMIS-PONCE, M.; ESPAÑA, K.; MACIAS, K. E. P.; FLORES, N. V. C.; ORLANDO, S. A.; PENAHERRERA, J. A. R.; CHEDRAUI, P.; ESCOBAR, S.; CHANGO, R. D. L.; RAMIREZ-MORÁN, C.; ESPINOZA-CAICEDO, J.; SÁNCHEZ-GILER, S.; LIMIA, C. M.; ALEMÁN, Y.; SOTO, Y.; KOURI, V.; CULASSO, A. C.; BADANO, I. Molecular epidemiology and phylogenetic analysis of human papillomavirus infection in women with cervical lesions and cancer from the coastal region of Ecuador. **Revista Argentina de Microbiología**, v. 50, n. 2, p. 136–146, abr. 2018. ISSN 03257541. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0325754117301372>>.

13 POTDAR, V.; VIPAT, V.; RAMDASI, A.; JADHAV, S.; PAWAR-PATIL, J.; WALIMBE, A.; PATIL, S.; CHOUDHURY, M.; SHASTRI, J.; AGRAWAL, S.; PAWAR, S.; LOLE, K.; ABRAHAM, P.; CHERIAN, S. Phylogenetic classification of the whole-genome sequences of SARS-CoV-2 from India & evolutionary trends. **Indian Journal of Medical Research**, v. 153, n. 1, p. 166, 2021. ISSN 0971-5916. Disponível em: <[https://journals.lww.com/ijmr/Fulltext/2021/01000/Phylogenetic\\_classification\\_of\\_the\\_whole\\_genome.14.aspx](https://journals.lww.com/ijmr/Fulltext/2021/01000/Phylogenetic_classification_of_the_whole_genome.14.aspx)>.

14 LICHTBLAU, D. Alignment-free genomic sequence comparison using FCGR and signal processing. **BMC Bioinformatics**, v. 20, n. 1, p. 742, dez. 2019. ISSN 1471-2105. Disponível em: <<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3330-3>>.

15 KIM, J.; CHEON, S.; AHN, I. NGS data vectorization, clustering, and finding key codons in SARS-CoV-2 variations. **BMC Bioinformatics**, v. 23, n. 1, p. 187, dez. 2022. ISSN 1471-2105. Disponível em: <<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-04718-7>>.

16 DIMITROV, K. M.; ABOLNIK, C.; AFONSO, C. L.; ALBINA, E.; BAHL, J.; BERG, M.; BRIAND, F.-X.; BROWN, I. H.; CHOI, K.-S.; CHVALA, I.; DIEL, D. G.; DURR, P. A.; FERREIRA, H. L.; FUSARO, A.; GIL, P.; GOUJGOULOVA, G. V.; GRUND, C.; HICKS, J. T.; JOANNIS, T. M.; TORCHETTI, M. K.; KOLOSOV, S.; LAMBRECHT, B.; LEWIS, N. S.; LIU, H.; LIU, H.; MCCULLOUGH, S.; MILLER, P. J.; MONNE, I.; MULLER, C. P.; MUNIR, M.; REISCHAK, D.; SABRA, M.; SAMAL, S. K.; ALMEIDA, R. Servan de; SHITTU, I.; SNOECK, C. J.; SUAREZ, D. L.; BORM, S. V.; WANG, Z.; WONG, F. Y. Updated unified phylogenetic classification system and revised nomenclature for Newcastle disease virus. **Infection, Genetics and Evolution**, v. 74, p. 103917, out. 2019. ISSN 15671348. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1567134819301388>>.

17 PEFFERS, K.; TUUNANEN, T.; ROTHENBERGER, M. A.; CHATTERJEE, S. A design science research methodology for information systems research. **Journal of management information systems**, Taylor & Francis, v. 24, n. 3, p. 45–77, 2007.



- 18 SILVA, D. D.; LOPES, E. L.; JUNIOR, S. S. B. Pesquisa Quantitativa: Elementos, Paradigmas e Definições. **Revista de Gestão e Secretariado**, v. 05, n. 01, p. 01–18, abr. 2014. ISSN 21789010, 21789010. Disponível em: <<http://www.revistagesec.org.br/ojs-2.4.5/index.php/secretariado/article/view/297>>.
- 19 ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. **Biologia Molecular da Célula**. [S.l.]: Artmed, 2017.
- 20 KNIPE, P. M. H. D. M. **Fields Virology**. 6. ed. [S.l.]: LIPPINCOTT WILLIAMS & WILKINS, 2022. Vol I and II. ISBN 9781451105636,1451105630,2013003842.
- 21 ZHU, N.; ZHANG, D.; WANG, W.; LI, X.; YANG, B.; SONG, J.; AL. et. A novel coronavirus from patients with pneumonia in china, 2019. **New England Journal of Medicine**, v. 382, n. 8, p. 727–733, 2020.
- 22 WU, F.; ZHAO, S.; YU, B.; CHEN, Y. M.; WANG, W.; SONG, Z. G.; AL. et. A new coronavirus associated with human respiratory disease in china. **Nature**, v. 579, n. 7798, p. 265–269, 2020.
- 23 LI, G.; FAN, Y.; LAI, Y.; HAN, T.; LI, Z.; ZHOU, P.; PAN, P.; WANG, W.; HU, D.; LIU, X.; ZHANG, Q.; WU, J. Coronavirus infections and immune responses. **J Med Virol**, v. 92, n. 4, p. 424–432, Apr 2020.
- 24 FELSENSTEIN, J. **Inferring Phylogenies**. 2. ed. [S.l.]: Sinauer Associates, 2004. ISBN 0878931775,9780878931774.
- 25 SWOFFORD, D. L.; OLSEN, G. J.; WADDELL, P. J.; HILLIS, D. M. Phylogenetic inference. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. (Ed.). **Molecular Systematics**. [S.l.]: Sinauer Associates, 1996. p. 407–514.