Universidade Federal do Rio Grande do Sul
Centro de Biotecnologia
Programa de Pós-Graduação em Biologia Celular e Molecular

Epidemiologia Genômica de SARS-CoV-2 no Estado do Rio Grande do Sul

Dissertação de Mestrado

Amanda de Menezes Mayer

Porto Alegre, 09 de setembro de 2022

Universidade Federal do Rio Grande do Sul
Centro de Biotecnologia
Programa de Pós-Graduação em Biologia Celular e Molecular

Epidemiologia Genômica de SARS-CoV-2 no Estado do Rio Grande do Sul

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da UFRGS como requisito parcial para obtenção do grau de Mestre.

Amanda de Menezes Mayer

Profa. Dra. Claudia Elizabeth Thompson – Orientadora
Profa. Dra. Lívia Kmetzsch Rosa e Silva – Coorientadora

Porto Alegre, 09 de setembro de 2022

Amanda de Menezes Mayer

Epidemiologia Genômica de SARS-CoV-2 no Estado do Rio Grande do Sul

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da UFRGS como requisito parcial para obtenção do grau de Mestre.

Data da avaliação: __/__/____

Conceito: _____

BANCA EXAMINADORA

_____

Profa. Dra. Claudia Elizabeth Thompson

Orientadora — UFCSPA

_____

Profa. Dra. Livia Kmetzsch Rosa e Silva

Orientadora — UFRGS

_____

Prof. Dr. Charley Christian Staats

Examinador Interno — UFRGS

_____

Prof. Dr. Luís Fernando Saraiva Macedo Timmers

Examinador Externo — UNIVATES

_____

Prof. Dr. Dennis Maletich Junqueira

Examinador Externo — UFSM

Porto Alegre 2022

Este trabalho foi desenvolvido na modalidade remota devido às restrições impostas pela pandemia de COVID-19 e contou com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) e Prefeitura Municipal de Esteio

Dedico este trabalho a todas as famílias brasileiras que enfrentaram obstáculos ou perdas devido à pandemia da COVID-19.

AGRADECIMENTOS

Gostaria de agradecer especialmente às seguintes pessoas, que foram imprescindíveis na realização desta dissertação:

Primeiramente aos meus pais, Márcia e Johni, por todo o amor e dedicação em minha criação e educação, por sempre apoiarem minhas escolhas e por todo o suporte dado para que eu realizasse esse trabalho.

Ao meu namorado, Patrik, que, com muito amor, me apoiou e acolheu durante esse processo, auxiliando na resolução de problemas e dando suporte às minhas crises. Obrigada por fazer com que as coisas se tornassem mais leves e felizes nesse período.

À minha gata, Amora, apenas por ser ela e ajudar a me distrair nos momentos de estresse.

À minha amiga, Francielle, por se fazer sempre presente e disponível para conversar, por me incentivar a não desistir, mesmo diante das dificuldades e pelo presente mais lindo e cheio de amor que é a Clara, minha afilhada.

A todos os colegas que, de alguma forma, estiveram presentes durante estes anos de mestrado e ajudaram este trabalho a se tornar real, sobretudo à Patrícia Ferrareze, Vinicius Franceschi e Gabriel Caldana por toda a parceria, pela ajuda durante a confecção dos trabalhos que fazem parte desta dissertação e na coleta de dados, pelas várias horas de reuniões e por terem me ensinado tantas coisas, sempre com muita paciência e dedicação.

À Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) pela concessão de minha bolsa de estudos, a qual tornou possível que eu me dedicasse integralmente à pesquisa durante este período.

Às minhas orientadoras, professoras Claudia e Lívia, pelos ensinamentos, auxílio prestado e disponibilidade durante a confecção desta dissertação.

Ao PPGBCM e a todos seus colaboradores e professores que sempre se mantiveram disponíveis e se esforçaram ao máximo para tornar o ensino à distância mais leve e produtivo, especialmente ao Prof. Dr. Hugo Verli que sempre prezou por auxiliar os alunos e manter o alto nível de ensino do programa, mesmo diante às dificuldades da pandemia e à Silvinha, por sempre estar disposta a amparar os alunos na resolução de problemas e por tratar as coisas com tanto carinho e afeição, mesmo à distância.

À Prefeitura de Esteio e todos aqueles que, de alguma forma, estiveram envolvidos na pesquisa GPS-COVID, que foi uma experiência bastante enriquecedora para minha vida pessoal e acadêmica, onde pude conhecer pessoas maravilhosas e obter conhecimentos que levarei para minha vida.

Ao Laboratório Central do Estado do Rio Grande do Sul (LACEN), pela disponibilização das amostras utilizadas neste trabalho e à BiomeHub pela realização do sequenciamento dessas amostras.

Aos membros da banca examinadora desta dissertação, Prof. Dr. Charley Staats, Prof. Dr. Luis Fernando Timmers e Prof. Dr. Dennis Junqueira, pela disponibilidade para a leitura deste trabalho, bem como pelas suas observações e recomendações direcionadas ao enriquecimento do estudo desenvolvido.

Aos alunos e professores da Escola Aldeia Lumiar e às colegas envolvidas no projeto em que desenvolvemos atividades didáticas. Este projeto foi um grande aprendizado, que me fez perceber a real paixão que tenho por ensinar e levarei sempre com muito carinho.

E, por fim, a todos os familiares e amigos que, de algum modo, se fizeram presentes e alegraram meus dias durante a realização deste trabalho.

“A ciência e a vida cotidiana não podem e não devem ser separadas.”

Rosalind Franklin

RESUMO

Com a emergência da pandemia da COVID-19, iniciada em 2020, diversos estudos vêm sendo conduzidos para entender a epidemiologia e a evolução do vírus SARS-CoV-2. É de extrema importância a análise desses trabalhos, em especial de suas metodologias, tal como realizado na revisão sistemática apresentada, que teve como foco avaliar estudos de prevalência da COVID-19. Desde a identificação do vírus SARS-CoV-2 como novo agente etiológico, diversas variantes já foram descritas e mutações definidoras de linhagem analisadas sob o ponto de vista funcional e evolutivo. Dentre elas, a mutação E484K, na proteína *spike*, se mostrou cada vez mais frequente e de importância nas linhagens onde se faz presente, principalmente por estar relacionada à indução de evasão imune. Este trabalho focou prioritariamente na análise genômica de amostras do estado do Rio Grande do Sul, em especial da cidade de Esteio, buscando entender o comportamento do SARS-CoV-2 dentro do território estadual. Dessa forma, obtivemos um panorama das mutações presentes nos genomas virais circulantes no estado, além de entendermos o comportamento evolutivo do vírus nesta localidade. Além disso, foi possível realizar testes de evolução molecular a fim de identificar possíveis evidências de pressão seletiva adaptativa e purificadora em sítios de proteínas estruturais do vírus, bem como analisar o efeito de diferentes mutações sobre a estabilidade molecular das mesmas. Por meio destes estudos, foi possível ampliar a compreensão da dinâmica genômica e evolutiva do vírus, bem como seus padrões de propagação pelo estado do Rio Grande do Sul.

ABSTRACT

With the emergency of COVID-19 pandemic starting in 2020, research has been conducted to understand the evolution and epidemiology of SARS-CoV-2 virus. It is of extreme importance the analysis of these studies, especially their methodologies, such as the presented systematic review which evaluated studies related to the prevalence of COVID-19. As the SARS-CoV-2 virus identification as a new etiological agent, several variants were described, and lineage-defining mutations were analyzed in evolutionary and structural point of views. Among them, the E484K mutation, in spike protein, became increasingly frequent and important in the strains where it is present, mainly because it is related to immune evasion. This work was focused on genomic analysis of samples of Rio Grande do Sul state, primarily from the city of Esteio, with the objective of understanding the behavior of SARS-CoV-2 inside the state territory. Thus, we got a wider view of the mutations present in the circulating viral genomes in the state, besides understanding the viral evolution behavior in that place. Furthermore, it was possible to realize molecular evolution analysis to identify the adaptive and purifying selective pressure evidence in structural proteins sites of the virus, as well analyze the mutational effects on its molecular stability. Through these studies, it was possible to expand the understanding of the viral genomic and evolutive dynamics, as well as their spreading patterns in the Rio Grande do Sul state.

LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES

cDNA	DNA complementar
COVID-19	Doença do coronavírus 2019
CTD	Domínio C-terminal
DNA	Ácido desoxirribonucleico
FEL	Probabilidade de efeitos fixos
FUBAR	Aplicação Bayesiana Rápida e Sem Limites
GISAID	Iniciativa Global de Compartilhamento de todos os Dados de Influenza
MAFFT	Alinhamento Múltiplo usando Fast Fourier Transform
MERS	Síndrome respiratória do Oriente Médio
NTD	Domínio N-terminal
NSP	Proteína não-estrutural
ORF	Quadro de leitura aberta
RBD	Domínio de ligação ao receptor
RNA	Ácido ribonucleico
RS	Rio Grande do Sul
RT-qPCR	Reação em cadeia da polimerase quantitativa com transcriptase reversa
S gene	Gene Spike
SARS	Síndrome respiratória aguda grave
SARS-CoV-2	Síndrome respiratória aguda grave do coronavírus 2
SLAC	Contagem de ancestrais de probabilidade única
SNPs	Polimorfismo de nucleotídeo único
UTR	Região não traduzida
VOCs	Variantes de preocupação
VOIs	Variantes de interesse
WHO	Organização Mundial da Saúde

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1. INTRODUÇÃO

Coronavírus são vírus de RNA envelopado que podem estar presentes em diversas espécies de animais e também em seres humanos, causando doenças respiratórias e entéricas (CUI; LI; SHI, 2019). Até então, sete coronavírus já foram identificados por causar doenças em humanos, sendo o SARS-CoV-2 o último deles, após os vírus SARS-CoV e MERS-CoV, que estiveram associados com epidemias em humanos, chamadas Síndrome Respiratória Aguda Grave (SARS) e Síndrome Respiratória do Oriente Médio (MERS), respectivamente. Destacam-se, como principais sintomas, dor de cabeça, febre e sintomas respiratórios que podem acabar progredindo para pneumonia (HU et al., 2015; ZHU et al., 2020).

O primeiro surto relacionado a SARS ocorreu em 2003, iniciando na província de Guangdong, na China e se espalhando para outros locais (KSIAZEK et al., 2003), seguido pela descoberta do SARS-CoV-2, agente etiológico da COVID-19, que foi anunciado pela Organização Mundial da Saúde (OMS) em dezembro de 2019, após um grupo de pacientes ter sido diagnosticado com pneumonia de origem desconhecida em Wuhan, província de Hubei, na China (PARK, 2020). Esta infecção tomou conta do mundo rapidamente, iniciando uma pandemia e levando a números desastrosos, os quais já ultrapassam 590 milhões de casos e 6 milhões de mortes pelo mundo (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2022).

Desde o início, os epicentros da pandemia se modificaram, iniciando em 2020 na Ásia, seguido da Europa e Américas. Com a evolução viral, novas variantes do SARS-CoV-2 surgiram. As principais variantes de preocupação (VOCs) até então são Alpha, Beta, Gamma, Delta e Ômicron (TELENTI; HODCROFT; ROBERTSON, 2022).

Descoberta em novembro de 2020, no Reino Unido, a variante Alpha (B.1.1.17) foi a primeira a ser descrita. Esta linhagem carrega diversas mutações no gene da proteína *spike* que aumentam a transmissibilidade viral (RAMBAUT et al., 2020). Ao mesmo tempo, o surgimento da variante Beta (B.1.351) ocorreu no continente Africano, acompanhada de algumas mutações de interesse na spike: D614G, D80A, D215G, E484K, N501Y e A701V (TEGALLY et al., 2020). Já em

dezembro de 2020, a variante denominada Gamma (P.1) foi detectada em Manaus, no norte do Brasil. Esta linhagem possui 17 mutações, sendo três destes importantes na proteína spike (E484K, K417T e N501Y) (FARIA et al., 2021). Além disso, estudos filogenéticos mostraram que a linhagem P.1 é claramente descendente da B.1.1.28, dividindo diversas mutações entre si (LAMARCA et al., 2021). As duas últimas VOCs descritas foram a Delta (B.1.617.2 e AY descendentes) e Ômicron (B.1.1.529 e BA descendentes). Em meados de novembro de 2020, a linhagem Delta foi detectada na Índia, demonstrando alta transmissibilidade e escape imunológico (DHAR et al., 2021). Relatada em novembro de 2021, a variante Ômicron, encontrada inicialmente na África, é a linhagem mais contrastante em comparação com as anteriores em virtude do seu alto número de mutações (HE et al., 2021).

Diversas mutações têm sido identificadas no genoma do SARS-CoV-2, levando à classificação em diferentes linhagens e distintas características clínicas da doença. Uma proporção significativa de mutações se encontra na região codificadora da proteína *spike* (BERRIO; GARTNER; WRAY, 2020). As mutações que resultam melhor adaptação viral ao meio são as que se mantêm no genoma como resultado da ação das forças evolutivas. No caso da proteína *spike*, aquelas que aumentam a transmissibilidade viral em virtude de sua função na ligação e acesso ao sistema imune do hospedeiro tendem a ser mantidas (PLANTE et al., 2021).

Uma das primeiras mutações descritas na proteína *spike* foi D614G, que está diretamente ligada à alta transmissibilidade viral e disseminação no organismo. Tal substituição se mantém presente em diversas das atuais VOCs (KORBER et al., 2020; GANGAVARAPU et al., 2022). Existente na maioria das linhagens - Alpha, Beta, Delta e Ômicron (GANGAVARAPU et al., 2022), a mutação N501Y também está relacionada à maior transmissão do vírus (LIU et al., 2022), tendo se tornado importante na evolução viral, juntamente da E484K, que está presente nas linhagens Beta e Gamma (GANGAVARAPU et al., 2022), e está associada ao escape imunológico, trazendo uma grande vantagem ao vírus (GREANEY et al., 2021; PAG et al., 2021).

A fim de auxiliar na resolução do quadro de pandemia, desde 2020, diversos

estudos têm sido conduzidos para o desenvolvimento de vacinas de uso emergencial (KRAMMER, 2020). Em 2021, a aplicação destes imunizantes foi instituída na população mundial, inicialmente mostrando uma boa eficácia no combate à doença (CHEN et al., 2021). Contudo, o desenvolvimento de maior resistência viral em virtude de novas mutações e linhagens têm levado à menor eficácia vacinal (WANG; CHEN; WEI, 2021). Estudos com a variante Ômicron indicaram alta resistência desta linhagem em relação aos anticorpos vacinais (PLANAS et al., 2021), o que demonstra que, embora as vacinas se mantenham eficazes principalmente contra desfechos graves da doença (ZHENG et al., 2022), o desenvolvimento de novos imunizantes de acordo com as novas linhagens emergentes se faz necessário para que seja obtida uma maior proteção contra a infecção (NOHYNEK; WILDER-SMITH, 2022).

Segundo a Organização Mundial da Saúde, o Brasil é o terceiro país com mais casos de COVID-19 confirmados, mais de 32 milhões, junto de 681 mil óbitos até agosto de 2022, deixando o país atrás apenas da Índia e Estados Unidos. Por ser um país com uma população extremamente heterogênea, o Brasil possui grupos de risco consideráveis para as formas mais graves da COVID-19, sejam estes por idade avançada ou comorbidades (como diabetes, hipertensão, doenças cardíacas ou respiratórias) (MARSON; ORTEGA, 2020).

Como mencionado anteriormente, a variante Gamma foi descrita inicialmente na cidade de Manaus, iniciando a segunda onda de infecções no Brasil, o que levou a um grande número de hospitalizações e, também em virtude da falta de suprimentos e insumos como falta de oxigênio nos hospitais, muitas mortes por COVID-19 ocorreram (FARIA et al., 2021; SABINO et al., 2021; SILVA; PENA, 2021).

Análises filodinâmicas demonstraram que esta variante circulou inicialmente do norte do país para as regiões centro-oeste e sul. Posteriormente, se espalhou para o restante do país (WOLF et al., 2022). A região sul do Brasil é composta por três estados: Paraná, Santa Catarina e Rio Grande do Sul (RS). De acordo com o Ministério da Saúde, o Rio Grande do Sul é o quarto estado brasileiro com mais casos de COVID-19, acumulando cerca de 2,6 milhões de casos confirmados até agosto de 2022, junto de 40 mil óbitos (SES/RS, 2022). As três cidades do estado

com maior número de casos são, em ordem decrescente, as cidades mais populosas: Porto Alegre, Caxias do Sul e Pelotas.

Porto Alegre é a capital do RS e compreende uma populosa Região Metropolitana, a maior do Brasil, formada por 34 cidades que concentram cerca de 38% da população do estado (SPGG/RS, 2020). Dentre estas cidades, a cidade de Esteio, localizada a 24km da capital, é a décima mais populosa e possui cerca de 83 mil habitantes (IBGE, 2021).

Desde o início da pandemia, Esteio acumulou pouco mais de 25 mil casos positivos de COVID-19 e 420 óbitos. A cidade já ultrapassou os 117 mil testes para detecção de SARS-CoV-2 (PREFEITURA DE ESTEIO, 2022) e se preocupou, desde o início da pandemia, em monitorar os casos e evitar a propagação do vírus por meio de medidas não-farmacológicas e testagem em massa. Um grande exemplo disto é o GPS-COVID, pesquisa executada de maio de 2020 até o final de 2021, com objetivo de mapear os perfis epidemiológico, genômico e clínico do SARS-CoV-2.

Considerando o exposto, este trabalho buscou realizar análises genômicas e evolutivas em genomas de SARS-CoV-2 do estado do Rio Grande do Sul e, dessa forma, caracterizar as particularidades em relação a suas mutações, padrões de dispersão e processos de evolução molecular envolvidos com o vírus.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Realizar análises genômica e evolutiva de amostras de SARS-CoV-2, com foco nas principais mutações e variantes, bem como sua propagação na cidade de Esteio e no estado do Rio Grande do Sul.

2.2. OBJETIVOS ESPECÍFICOS

- Realizar identificação e análise das mutações nos genomas de SARS-CoV-2;
- Analisar a filogenômica de SARS-CoV-2 com o intuito de identificar seus clados e agrupamentos;
- Realizar estudos de evolução molecular, a fim de identificar a ação de seleção positiva no genoma SARS-CoV-2;
- Realizar estudos sobre estabilidade molecular das proteínas estruturais e sua relação com a evolução do vírus.

3. CAPÍTULO I

O manuscrito que compõe o presente capítulo, intitulado “Genomic characterization and molecular evolution of SARS-CoV-2 in Rio Grande do Sul State, Brazil” teve por objetivo descrever a dinâmica molecular e características genômicas do SARS-CoV-2 circulante no município de Esteio (RS) e compará-las com o estado do Rio Grande do Sul. Para isto, foram utilizadas 12 genomas provenientes de pacientes hospitalizados por COVID-19 no Hospital São Camilo, na cidade de Esteio, no período de abril a junho de 2021, juntamente de 2.227 genomas do estado do Rio Grande do Sul do período de março de 2020 a maio de 2022 que foram disponibilizados pela plataforma GISAID. Foram realizadas análises filogenômicas e de evolução molecular.

Este artigo encontra-se publicado na revista Virology (https://www.sciencedirect.com/journal/virology) que possui Fator de Impacto (JCR 2022) = 3.7 e Qualis/CAPES = A2 e todas as análises nele descritas, bem como sua redação, foram realizadas pela aluna Amanda de Menezes Mayer, sendo os demais autores colaboradores no processo de escrita e de análises, assim como sua orientação. O manuscrito, juntamente de seus materiais suplementares se encontram disponíveis na íntegra *(Open Access)* e podem ser acessados por meio do seguinte endereço: https://doi.org/10.1016/j.virol.2023.03.005.

Os genomas sequenciados do município de Esteio (RS), utilizados neste estudo, atualmente estão sendo organizados para submissão na plataforma GISAID.

Genomic characterization and molecular evolution of SARS-CoV-2 in Rio Grande do Sul State, Brazil

Amanda de Menezes Mayer[1], Patrícia Aline Gröhs Ferrareze[2], Luiz Felipe Valter de Oliveira[3], Tatiana Schäffer Gregianini[4], Gabriel Dickin Caldana[2], Ricardo Ariel Zimerman, Lívia Kmetzsch[1], Claudia Elizabeth Thompson[1, 2, 5, *]

[1] Center of Biotechnology, Graduate Program in Cell and Molecular Biology (PPGBCM), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

[2] Graduate Program in Health Sciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil

[3] BiomeHub, Florianópolis, Brazil.

[4] Laboratório Central de Saúde Pública da Secretaria de Saúde do Estado do Rio Grande do Sul (LACEN/CEVS/SES-RS)

[5] Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, Brazil

E-mail: cthompson@ufcspa.edu.br, thompson.ufcspa@gmail.com

*** Corresponding author**

Address for correspondence:

Claudia Elizabeth Thompson

Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), 245/200C Sarmento Leite St, Porto Alegre, RS, Brazil. ZIP code: 90050-

170. Phone: +55 (51) 3303 8889. E-mail: cthompson@ufcspa.edu.br, thompson.ufcspa@gmail.com

Running title: SARS-CoV-2 Genomics in Rio Grande do Sul State

Keywords: SARS-CoV-2, Genomics, COVID-19, molecular evolution, phylogenomics

Abstract

The SARS-CoV-2 is the virus responsible for the COVID-19 pandemic and is plaguing the world since the end of 2019. Different lineages have been discovered ever since and the Gamma lineage, which started the second wave of infections, was first described in Brazil, one of the most affected countries by pandemic. Understanding the viral genome and how the virus behaves is essential to contain its propagation and to the development of medications and vaccines. Therefore, this study analyzed SARS-CoV-2 sequenced genomes from Esteio city in Rio Grande do Sul, Southern Brazil. We also comparatively analyzed genomes of the two first years of the pandemic from Rio Grande do Sul state for understanding their genomic and evolutionary patterns. The phylogenomic analysis showed that the lineages keep a high pattern of monophyletic groups in the state. However, some Esteio genomes are related with sequences from other countries, indicating viral importation events. Molecular evolution analyzes identified several sites under adaptive selection in the structural proteins of the virus and using these results the stability of the protein structures was evaluated. This study provides a better understanding of the viral genomic and evolutionary arrangement in the state of Rio Grande do Sul, Brazil.

Introduction

After the first outbreak of COVID-19 (Coronavirus disease 2019) occured in Wuhan, Hubei Province, China in December 2019 (Zhu et al., 2020), the new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spread around the world. Starting a world pandemic, declared by the World Health Organization (WHO) in March 2020, COVID-19 already exceeds 570 million cases and 6.39 million deaths until July 2022 (Dong et al., 2020). Currently (July, 2022), Brazil is the third country most affected by SARS-CoV-2, reaching the mark of 33.7 million confirmed cases and more than 677 thousands deaths (Dong et al., 2020). Of these, 7.8% of the cases and 5.97% of the deaths are from Rio Grande do Sul (RS), the southernmost state of Brazilian territory and the fourth state in the ranking of COVID-19 cases.

Among several lineages, along these two years of pandemic, various Variants of Concern (VOCs), such as Alpha, Beta, Gamma, Delta, and Omicron, carrying signature aminoacid substitutions (especially in the spike protein) have been circulating in RS state (Gularte et al., 2022; Wink et al., 2022). The sublineage P.1.2, a Gamma-like variant, was found to have probably arisen in this region, in February, 2021, and posteriorly spread across the country (Franceschi et al., 2021). As in Brazil, an increasing number of cases and deaths by Gamma (P.1 lineage) became evident in RS, in the beginning of 2021, causing a second COVID-19 wave. The P.1 lineage harbors mutations in the Spike's receptor-binding domain (RBD) such as E484K, K417T, and N501Y, which promote evasion from antibody neutralization elicited by infection or vaccination (R. E. Chen et al., 2021; Chakraborty, 2022).

In December, 2020, the Delta variant was described initially in India. This variant is highly transmissible and spreads easily, causing new waves of infection around the world by the middle 2021 (Shiehzadegan et al., 2021). However, despite this lineage rapidly becoming dominant in Brazil (including RS state), in July/August, 2020, there was not reported a concurrent increase in reported cases or deaths (Giovanetti et al., 2022). By the end of 2021, the number of cases of COVID-19 were progressively decreasing, until the emergence of the Omicron variant, confirmed in November 2021, in South Africa (Wang & Powell, 2021). This new variant is a concern due the mutations on RBD and cleavage sites that suggest higher transmissibility, up to three times more contagious than Delta variant (J. Chen et al., 2022).

Despite all VOC characterizations mostly focused on spike mutations, structural proteins E (Envelope), M (Membrane), and N (Nucleocapsid) are functionally important to virus assembly and pathogenesis (Yadav et al., 2021). The N protein has been associated to the promotion of inflammatory processes by activation of COX-2, to interaction with p42 proteasome in order to avoid the degradation of viral proteins and to inhibition of IFN-I in immune response (Satarker & Nampoothiri, 2020). Moreover, M protein is known to inhibit NFκB, to reduce levels of COX-2, to activate IFN-β and to interact with PDK1/PKB proteins, leading to cell death or apoptosis.

In this way, this study aims to perform genomic sequencing and characterization of the SARS-CoV-2 genomes from RS state as well to identify selection traits in E, M, and N protein sites, elucidating the molecular evolution

processes that drive the diversification or conservation of the structural proteins from SARS-CoV-2.

Methods

Ethics approval

Ethical approval was obtained from Comitê de Ética em Pesquisa em Seres Humanos da Universidade Federal de Ciências da Saúde de Porto Alegre (CEP - UFCSPA) under process number CAEE 39247920.0.0000.5345.

Sample collection and clinical testing

Respiratory secretion were analyzed by Laboratório Central de Saúde Pública do Estado do Rio Grande do Sul (LACEN) (Porto Alegre, Rio Grande do Sul, Brazil) using RT-qPCR AllPlex SARS-CoV-2 assays Seegene Inc. Seoul, Republic of Korea with primers and probes targeting the RNA dependent RNA Polymerase (RdRP) Nucleocapsid (N) and Envelope (E) genes as recommended by the World Health Organization, with remnant samples stored at -20ºC. For the sequencing protocol, positive samples in the first RT-qPCR between April 09, 2021 to June 29, 2021, were selected and submitted to a second RT-qPCR, which was performed by BiomeHub (Florianópolis, Santa Catarina, Brazil), with a charite-berlin protocol. Samples with quantification cycle (Cq) up to 30 for at least one primer were selected for SARS-CoV-2 genome sequencing and assembly by the BiomeHub laboratory. In total, 12 patients who tested positive for SARS-CoV-2 RT-qPCR were included in the study.

SARS-CoV-2 genome sequencing and assembly

Total RNAs were prepared according to a reference protocol (Eden & Sim, 2020), with cDNA synthesized with SuperScript IV (Invitrogen) and DNA amplified with Platinum Taq High Fidelity (Invitrogen). The library preparation was performed with Nextera Flex (Illumina) and quantification was performed with Picogreen and Collibri Library Quantification Kit (Invitrogen). The genome sequencing was generated on Illumina MiSeq Platform by 150x150 runs with 500xSARS-CoV-2 coverage (50-100 mil reads/per sample).

For the genome assembly (BiomeHub in-house script), the adapters removal and read trimming for 150 nt read sequences were performed by fastqtools.py. The alignment of the sequenced reads to the reference SARS-CoV-2 genome (GenBank ID: NC_045512.2) was performed by Bowtie v2.4.2 (Langmead & Salzberg, 2012) with additional parameters as end-to-end and very-sensitive. The analyses of the sequencing coverage and depth were generated by samtools v1.11 (Li et al., 2009) with minimum base quality per base (Q) ≥ 30. Finally, the consensus sequence for each SARS-CoV-2 genome was generated by a bcftools pipeline (Li, 2011), including the commands mpileup (parameters: Q ≥ 30; depth (d) ≤ 1,000), filter (parameters: DP>50), and consensus.

SARS-CoV-2 genomes and data retrieval

In order to compare our samples to the samples from the state, we gathered all sequences uploaded on GISAID (Elbe & Buckland-Merrett, 2017) in the same time span of our collection. On May 27th, 2022, the GISAID search yielded 2227 genomes after applying the filters: Location: South America / Brazil / Rio Grande do Sul; Clade: all; Complete genomes; High Coverage Selected.

SARS-CoV-2 mutations and lineages

SNPs and insertions/deletions in each sample were identified by the variant calling pipeline (https://github.com/tseemann/snippy), which uses FreeBayes and snpEff to call, annotate and predict variant effects on genes and proteins. The genomes were aligned with MAFFT and the extraction of SNPs and gaps from the sequences, in relation to the reference, with msastats.py script. The reference sequence comes from the GenBank RefSeq (NC_045512.2), isolated and sequenced from an initial case from Wuhan, China, in 2019. The strains were identified using the dynamic nomenclature implemented in Pangolin (Rambaut, 2020) (https://github.com/cov-lineages/pangolin) and global clades and mutations using Nextstrain from Nextclade (https://clades.nextstrain.org/).

Phylogenomic analyses

For the global phylogenomics, a search was performed by Audacity *Instant* on the GISAID database (Elbe & Buckland-Merrett, 2017) to find closely related sequences to the sequenced genomes from this study (up to June 23, 2022). The resulting genome set was aligned with the MAFFT v.7 web server (Katoh et al., 2017). The trimming of 5' and 3' UTRs was performed with UGENE (Okonechnikov et al., 2012). The evolutionary model and phylogenomic tree inference was performed made with by IQ-TREE software (Nguyen et al., 2014) with addition of a Shimodaira-Hasegawa-like approximate likelihood ratio test of 1,000 replicates (Guindon et al., 2010) and an approximate Bayes test (Anisimova et al., 2011). Figtree software (http://tree.bio.ed.ac.uk/software/figtree/) was used to inspect and visualize the phylogenetic tree.

For the local phylogenomic analyses, genome sequences from Rio Grande do Sul state, previously downloaded from GISAID database, were aligned using MAFFT v.7 web server (referência). The trimming of 5' and 3' UTRs was performed with UGENE (Okonechnikov et al., 2012). The evolutionary model and phylogenomic tree inference was performed by IQ-TREE software (Nguyen et al., 2014) with addition of a Shimodaira-Hasegawa-like approximate likelihood ratio test of 1,000 replicates (Guindon et al., 2010) and an approximate Bayes test (Anisimova et al., 2011). Figtree software (http://tree.bio.ed.ac.uk/software/figtree/) was used to inspect and visualize the phylogenetic tree.

Phylogenetics and Molecular Evolution of SARS-CoV-2 Structural Proteins

In order to infer the phylogenetic patterns of structural proteins E, M and N, genomic alignment coordinates related to these sequences were exported, according to SARS-CoV-2 reference genome (NC_045512.2). Sequences with nucleotide insertions altering the reading frame were excluded from the analysis. For each gene sequence alignment, the evolutionary model and phylogenetic tree were inferred according to the previously described parameters from phylogenomic analysis.

Molecular evolution tests were performed with the HyPhy package (Pond et al., 2004). The methods FUBAR (Murrell et al., 2013), FEL (Kosakovsky Pond & Frost, 2005), and SLAC (Kosakovsky Pond & Frost, 2005) were implemented in order to evaluate potential sites under adaptive (pervasive) and purifying selection.

Molecular stability of structural proteins E, M and N

The estimation of molecular stability of structural proteins was performed by DynaMut web server (Rodrigues et al., 2018) using the experimentally resolved crystal structures (a) 7K3G - relative to transmembrane domain of protein E; (b) 8CTK - relative to protein M; (c) 7VNU - relative to N-terminal domain of protein N; and (d) 6ZCO - relative to C-terminal domain of protein N, from Protein Data Bank (https://www.rcsb.org/). The selection of tested amino acid mutations was defined according to the sites detected under positive selection by the molecular evolution tests.

RESULTS

Genomic epidemiology

Twelve samples were collected from hospitalized patients from the municipality of Esteio, RS state, between April 9th and June 29th, 2021. The mean cycle threshold (Ct) value for the first RT-qPCR conducted at Laboratório Central de Saúde Pública do Estado do Rio Grande do Sul (LACEN) was 23.83 cycles (median: 23.00; IQR: 4.5 cycles).

The sequence coverage for the twelve sequenced genomes ranged between 84.92 and 99.76% (mean: 98.26%) of the 29,903 bp of NC_045512.2 reference genome. The mean of sequencing depth was calculated to 292.23x, with a variation between 53.15 and 542.28x. Leastwise 48.81% of the sequence nucleotides accomplish a coverage depth ≥51x (max: 98.27%, mean: 87.68%) (Supplementary File 1). According to Pango lineage assignment, 10 sequenced samples belong to P.1 lineage and 2 are from P.1.17 sublineage, both from Gamma variant clade.

Genomic characterization

Sixty-eight different nucleotide substitutions were found in the twelve genomes from this study, being 36 of these non-synonymous polymorphisms. The non-synonymous mutations E1264D (ORF1b), D614G and V1176F (S) were predominantly found, identified in the 12 sequences.

The most common substitutions on S gene (in at least 75% of the sequences) were K417T and T1027I (in 11 genomes), and H655Y (in 9 genomes). Only two samples carry substitutions E484K and N501Y, being both mutations in the same samples. The mutations E611D and S689I are present in one sample each.

A few mutations found in our samples are described for the first time in RS (Table 1). Most of them were already identified in the world and occur in sequences from different VOCs such as Alpha, Beta, Gamma, Delta and Omicron. Substitution D1208A (NSP3), specifically, was also found for the first time in Brazil. The non-synonymous mutation I136V, localized in the NSP12b gene wasn't described on GISAID, being ours the first report about this mutation (Table 1).

Table 1. List of mutations firstly described in Rio Grande do Sul in sequenced genomes from this study.

Mutation	Occurrences in the world	First occurrence in world	Occurrences in Brazil	First occurrence in Brazil	Date of our sample	VOCs
A72S	7,462	2020	85	2021-02-03	2021-05-01	A, B, G, D, O
P1103L	17,219	2020	428	2020-06-16	2021-06-14	A, B, G, D, O
D1208A	271	2020-03-13	not described	-	2021-06-14	A, G, D, O
D144G	198	2020	2	2021-05-31	2021-06-14	A, G, D, O

L325F	1,255	2020	15	2021-01-27	2021-06-14	A, B, G, D, O
I136V	not described	-	not described	-	2021-05-08	G

The table describes the mutations and the number of occurrences of each in the world and Brazil, as well as his first occurrences in these locations. We also can see in the table the VOC lineages where each mutation can be found (Access on GISAID: June 7, 2022). A: Alpha; B: Beta; G: Gamma; D: Delta; O: Omicron.

Genomic characterization from Rio Grande do Sul state SARS-CoV-2 sequences

A number of 2,227 sequences were downloaded from the GISAID platform ranging from March, 2020 up to May, 2022 (26 months). The lineage frequencies in the whole period are presented in Supplemental file 1. One sample yielded “Unassigned” lineage, but analyzing the mutations it is possible to infer that they are more likely to belong to the Omicron group. The genome sequencing count per month in RS state can be visualized in Figure 1. There was a downward tendency in sequences count throughout the time with a mean of 82 sequenced samples per month. Sadly, the total number of sequences in the state is low compared to the number of cases, representing around 0.09% of them.

As shown in Figure 2, from January until July 2021, the P.1 lineage was prevalent in the RS genomes; that matches with the lineages of our samples, which are also P.1 and were collected between April and June 2021, during the GAMMA-related second wave of COVID-19 in the state. Thirty-two percent of SARS-CoV-2 genomes obtained on the GISAID database belong to P.1 lineage. The P.1.2, P.2 and P.7 lineages achieved higher frequencies in the beginning of the year, but P.2 remained with a low number of cases until July. The Delta lineages arrived in the state in June 2021 and became prevalent from August up to the end of the year.

Delta lineage AY.99.2 was accountable for 17% of the sequenced genomes from RS state in 2021.

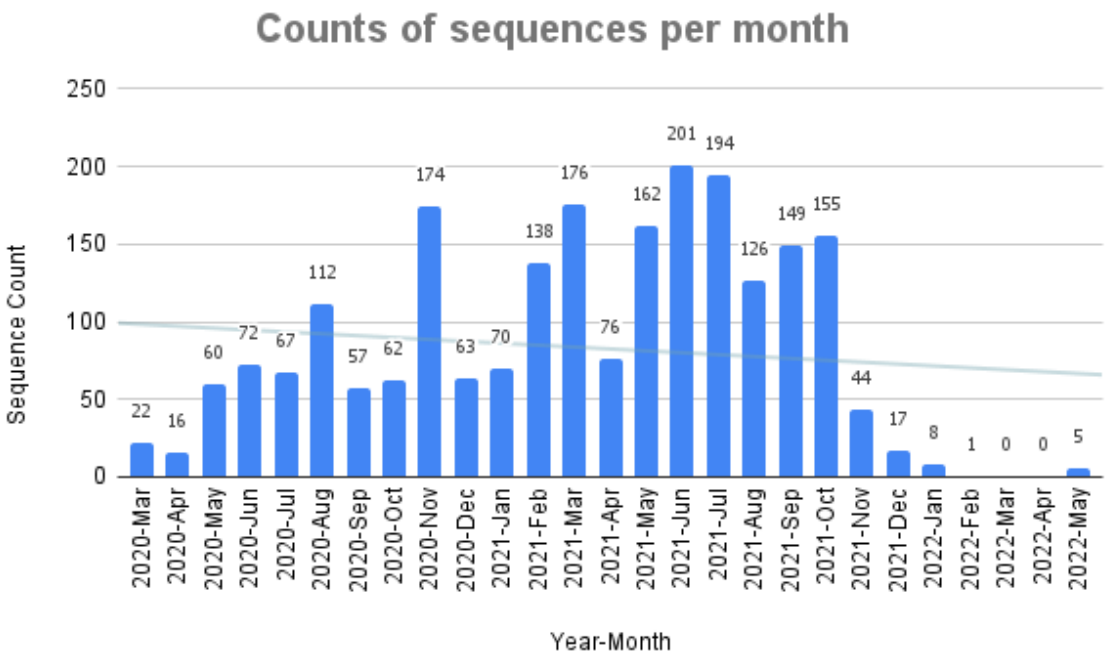


Figure 1. Counts of sequences per month in Rio Grande do Sul, between March 2020 and May 2022. The light blue line shows the sequences had a downward tendency in their counts, throughout the time with a mean of 82 sequenced samples per month.

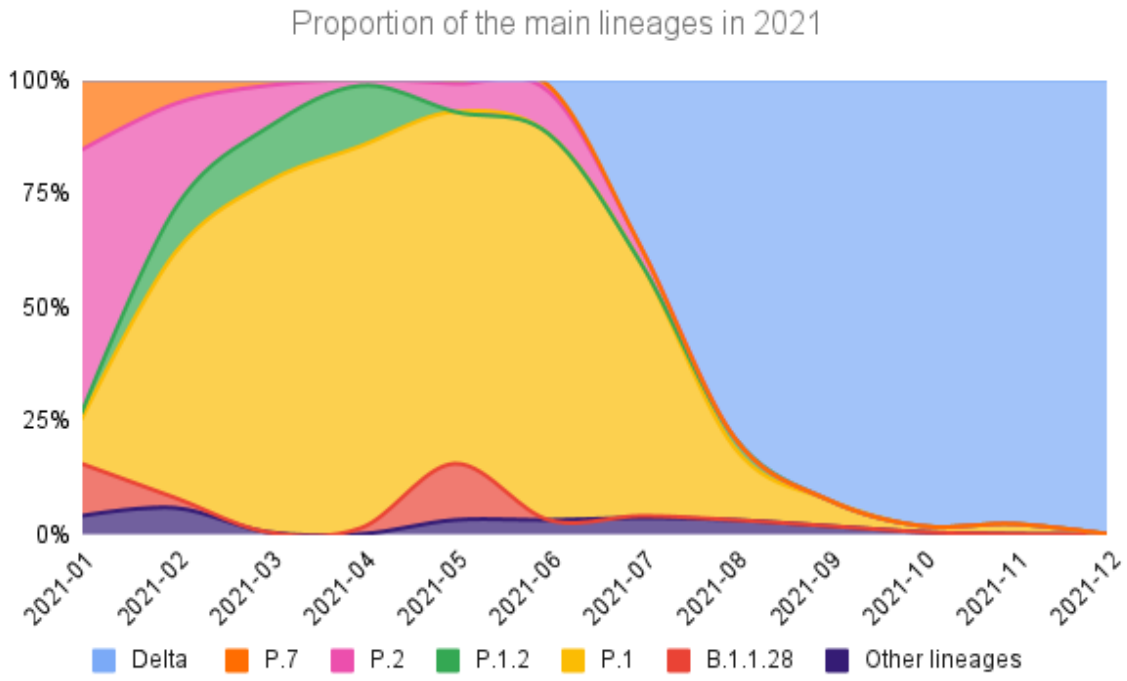


Figure 2. Lineage distribution in Rio Grande do Sul state at 2021. **Delta comprises B.1.617.2 derivative lineages: AY.20, AY.34, AY.4, AY.43, AY.43.1, AY.45, AY.46.3, AY.9.2, AY.99, AY.99.2, AY.100, AY.101 and AY.103. Other lineages comprises P.1.7, P.1.8, P.1.10, P.1.14, P.1.17, B, B.1, B.1.1.161, B.1.1.7, B.1.1.33, B.1.332, B.1.575 and C.37.**

About the genome set from RS state, the missense variant NSP12:P314L (ORF1b) was the most prevalent, found in 99.10% (n = 2,207) of samples (Figure 3). The highly frequent mutations (present in at least 90% of the genomes) also include the extragenic substitution C241T (n = 2,153), the synonymous mutation NSP3:F106F (n = 2,103) and the non-synonymous mutation S:D614G (n = 2,061). Other non-synonymous mutations such as N:R203K/G204R (n = 1,640) and S:V1176F (n = 1,311) were found in > 50% of samples.

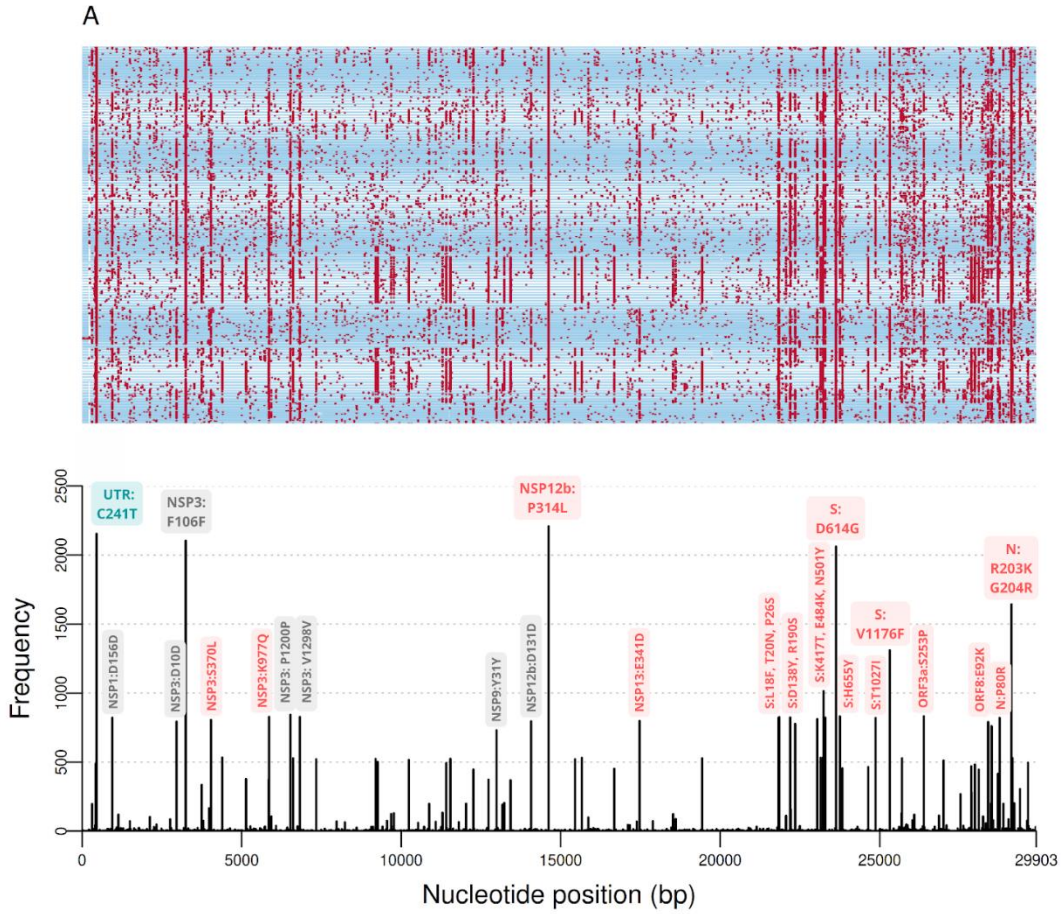


Figure 3. Amino acid and nucleotide substitutions associated with SARS-CoV-2 genomes from RS state. Synonymous and non-synonymous mutations are labeled with the amino acid residue. Substitutions in extragenic positions were labeled with the nucleotide alteration. Mutations occurring in more than 30% of the samples were labeled in different colors: red (non-synonymous mutations), gray (synonymous mutations), blue (extragenic mutations), green (nucleotide insertions and amino acid deletions). Mutation N:S201S (n = 817) and insertion CAAA in position 28268 (N = 755) are not labeled in the plot.

Global Phylogenomics

In order to locate these sequenced samples in SARS-CoV-2 global dataset, the AudacityInstant tool from GISAID database was used to identify genetically

related genomes. Four sequences could not be related to other sequences in the database, probably due to low sequencing quality. Considering the most related genomes (Table 2), four sequences were associated with samples from Brazil (one of them from Rio Grande do Sul state) and the other four were more closely related to genomes from Chile, Mexico, USA and Canada.

Table 2. Closest related SARS-CoV-2 sequences from GISAID to the sequenced genomes from this study according to AudacityInstant search.

Sequence	Closest related genome				
	Distance	Match quality	Location	Collection date	Lineage
RS-44473	No related genomes found				
RS-44474	No related genomes found				
RS-44475	4	0.910	Brazil / São Paulo	2021-11-22	P.1
RS-44476	3	0.920	Canada	2021-05-26	P.1.17
RS-44477	No related genomes found				
RS-44478	4	0.900	Chile	2021-08-13	B.1.1
RS-44479	8	0.937	Brazil / Rio de Janeiro	2021-02-10	P.1
RS-44480	2	0.905	Brazil / São Paulo	2021-06-07	P.1
RS-44481	1	0.971	Brazil / Rio Grande do Sul / Guaíba	2021-06-24	P.1
RS-44482	5	0.901	USA	2021-03-17	P.1.13
RS-44483	No related genomes found				
RS-44484	4	0.903	Mexico	2021-06-05	P.1.17

Besides the most closely related genomes, other 424 unique sequences (638 genomes in total) were recovered as being related to the sequenced genomes from this study with a genetic distance of 9 or less according to AudacityInstant parameters (Figure 4). Sequences RS-44475, RS-44478, RS-44479 and RS-44481

were mostly associated with Brazilian sequences (47.4 up to 89% of the retrieved genomes). RS-44476, RS-44480 and RS-44484 were predominantly related to Mexico (50%) and USA (30.3 and 27.5% of the genomes), respectively.

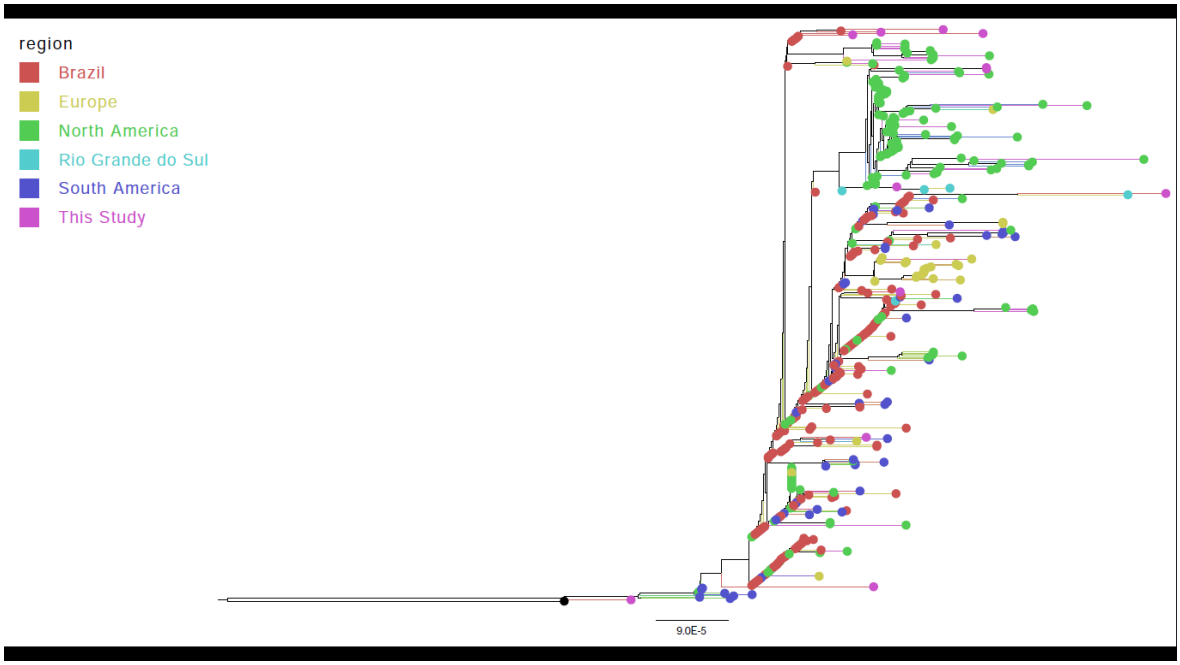


Figure 4. Global Maximum likelihood phylogenomic analysis with SARS-CoV-2 genomes closely related to the 12 sequenced genomes from this study.

Phylogenomic analysis of SARS-CoV-2 from Rio Grande do Sul state

The phylogenomic analysis of the SARS-CoV-2 genomes from Rio Grande do Sul state showed the formation of multiple monophyletic groups for the main VOCs (Figure 5). Alpha (B.1.1.7), Gamma (P.1 and derivative lineages), Delta (B.1.617.2 and derivative lineages) and Omicron (BA.2) clades were validated by SH-aLRT and aBayes tests with at least 97% of branch support (100/1, 99.9/1, 97.1/1, and 100/1 for Alpha, Gamma, Delta and Omicron, respectively). For other lineages and former VOIs, such as B.1.91, P.7 and Zeta (P.2) it was also observed the clustering in monophyletic groups (95.6/1, 97.1/1, and 99/1 of statistical support

for B.1.91, P.7 and P.2, respectively). Despite the presence of B.1.1.28 genomes at the root of the Gamma clade, the known ancestor-descendant relationship between them did not evidence the formation of a monophyletic group since B.1.1.28 sequences are dispersed along the phylogeny. Interestingly, a clade with Alpha and Omicron genomes was formed with 86.2/0.996 of branch support.

As expected, all 12 genomes sequenced by this study clustered in the Gamma clade, which also presented subclades related to P.1 sublineages. P.1.2 (94/1 for SH-aLRT/aBayes), P.1.17 (85.3/0.997 for SH-aLRT/aBayes, including two genomes from this study at the basal branch), and P.1.7 (88.7/1 for SH-aLRT/aBayes) were found to form monophyletic groups. In the Delta group, sublineages AY.101 and AY.9.2 were supported as subclades by the statistical tests (96.3/1 and 100/1 for SH-aLRT/aBayes, respectively).

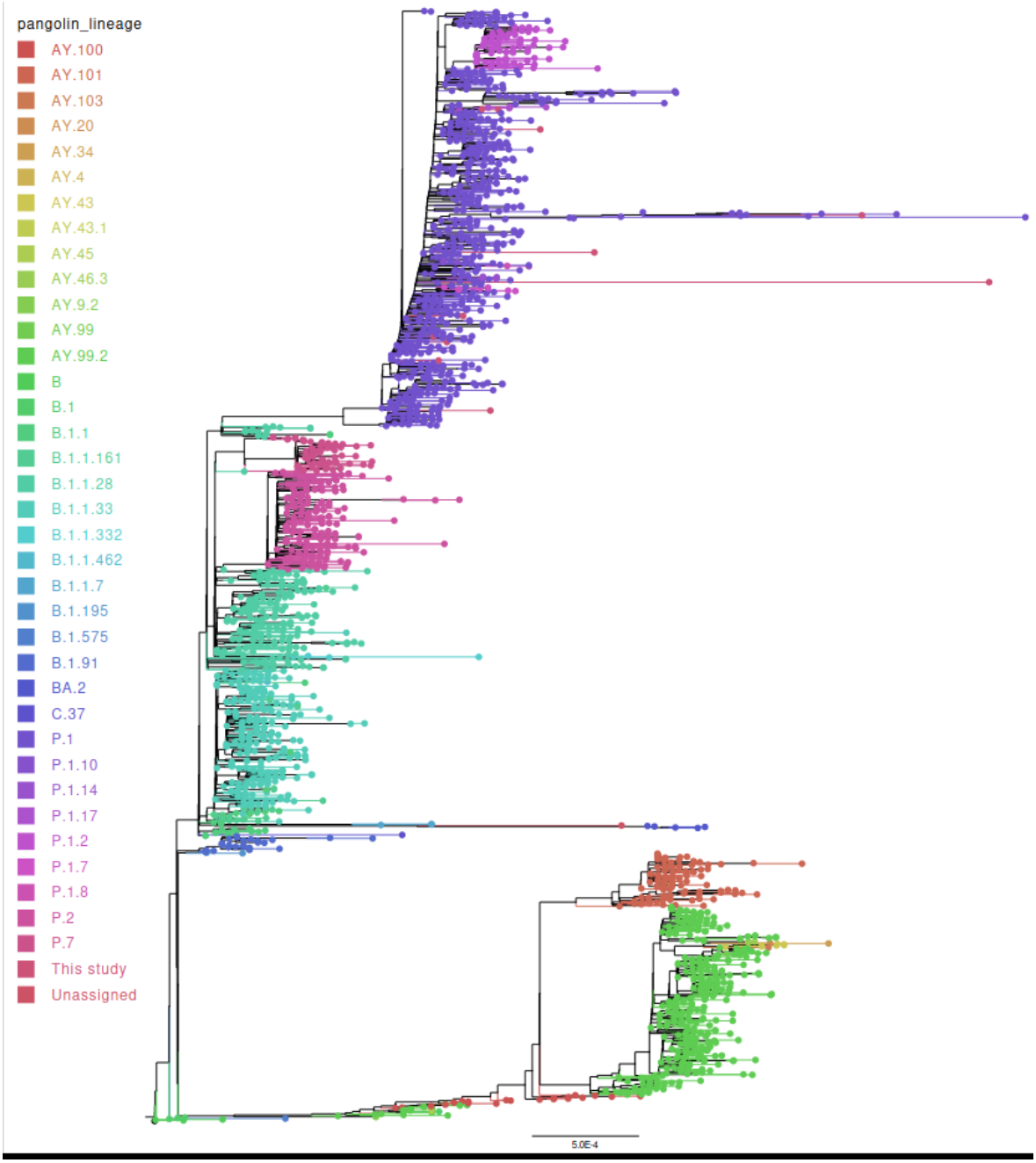


Figure 5. Maximum likelihood phylogenomic analysis of SARS-CoV-2 genomes from Rio Grande do Sul state.

Phylogenetics and Molecular Evolution of SARS-CoV-2 Structural Proteins

The molecular evolution analysis aimed to identify positively and negatively selected sites from SARS-CoV-2 structural proteins in the genome dataset from Rio

Grande do Sul state. In this way, E, M, and N proteins were tested with HyPhy FUBAR, FEL and SLAC methods (Tables 3 - 5 and Supplementary File 2).

E protein

Despite ten sites being identified by FEL and SLAC for E protein, none of them was found by both methods (Supplementary File 2). Thus, sites 7, 9, 23, 30, 55, 61, 66, 68, 72 and 73 were found to be under diversifying selective pressure.

M protein

Twenty-nine sites were identified to be under adaptive selection by FEL and/or SLAC in M protein (Supplementary File 2). Three of them were identified by both methods (Table 3). Other three sites (53, 109 and 162) were found to be under purifying selection.

Table 3. Protein M sites subjected to positive and negative selection according to HyPhy methods.

Codon	FUBAR			FEL				SLAC		
	Alpha	Beta	Post. prob.	Alpha	Beta	LRT	Prob.	dS	dN	Prob
63	-	-	-	0.000	3.395	3.389	0.0656	22.952	50.757	0.000
86	-	-	-	0.000	3.387	2.895	0.0889	0.000	49.412	0.000
125	-	-	-	0.000	5.193	3.556	0.0593	0.000	35.069	0.000

Post. prob.: Posterior probability. Prob.: Probability

N protein

One hundred and fifteen sites were identified to be under selective pressure by FUBAR, FEL and/or SLAC (Supplementary File 2). Of these, fifteen sites

presented evidence of purifying selective pressure. Twenty-two sites were identified by two methods, at least (Table 4).

Table 4. Protein N sites subjected to positive and negative selection according to HyPhy methods.

Codon	FUBAR			FEL				SLAC		
	Alpha	Beta	Post. prob.	Alpha	Beta	LRT	Prob.	dS	dN	Prob
9	1.162	6.689	0.9171	0.000	11.856	8.821	0.0030	-	-	-
34	0.627	6.934	0.9840	0.000	2.294	3.605	0.0576	-	-	-
63	0.626	6.643	0.9772	0.000	8.489	11.415	0.0007	0.000	30.509	0.000
103	-	-	-	0.000	2.296	2.899	0.0886	0.000	17.188	0.000
110	-	-	-	2.022	0.000	2.844	0.0917	2.391	0.000	0.078
151	0.595	7.120	0.9879	0.000	5.180	7.784	0.0053	37.283	81.437	0.000
182	0.610	4.020	0.9256	-	-	-	-	46.782	101.306	0.000
185	-	-	-	0.000	2.593	3.879	0.0489	11.101	27.546	0.003
202	-	-	-	0.000	4.482	6.323	0.0119	0.000	17.927	0.000
204	-	-	-	0.000	4.375	2.854	0.0911	31.486	64.818	0.000
208	0.604	4.019	0.9266	0.000	2.751	3.977	0.0461	33.854	74.294	0.000
215	0.620	3.320	0.9035	0.000	4.566	7.028	0.0080	1.095	5.805	0.058
238	0.613	5.716	0.9707	0.000	5.145	7.604	0.0058	28.146	100.807	0.000
274	-	-	-	3.080	0.000	3.973	0.0462	2.392	0.000	0.078
289	1.190	6.874	0.9166	0.000	6.378	4.554	0.0328	176.613	40.051	0.000
362	1.235	7.839	0.9224	0.000	7.674	4.785	0.0287	0.000	10.862	0.000
366	1.234	7.818	0.9223	0.000	7.672	4.767	0.0290	0.000	11.841	0.000
368	-	-	-	0.435	8.488	7.846	0.0051	11.010	26.463	0.004
383	-	-	-	2.259	8.637	2.848	0.0915	113.407	91.536	0.054
391	0.602	5.675	0.9571	0.000	7.591	9.092	0.0026	-	-	-
401	-	-	-	0.000	2.543	3.029	0.0818	0.000	28.324	0.000
413	-	-	-	0.000	2.462	2.990	0.0838	0.000	27.394	0.000

Post. prob.: Posterior probability. Prob.: Probability. Sites subjected to negative selection are highlighted in gray.

Molecular stability of structural proteins E, M and N

The program DynaMut was performed to estimate the molecular stability of the SARS-CoV-2 structural proteins E, M and N with mutated residues at sites previously identified under positive selection (Table 6). In this way, sites recognized by two or more methods in HyPhy tests had their amino acid substitutions evaluated at molecular level with publicly available crystallographic structures from PDB database (Figure 6). Since protein E had no sites repeatedly identified, all sites were considered for analysis.

Differently from spike protein, proteins E, M and N are less represented in experimentally resolved structures from PDB. Thus, structures: (a) 7K3G - relative to transmembrane domain of protein E; (b) 8CTK - relative to protein M; (c) 7VNU - relative to N-terminal domain of protein N; and (d) 6ZCO - relative to C-terminal domain of protein N, were selected to perform these estimatives.

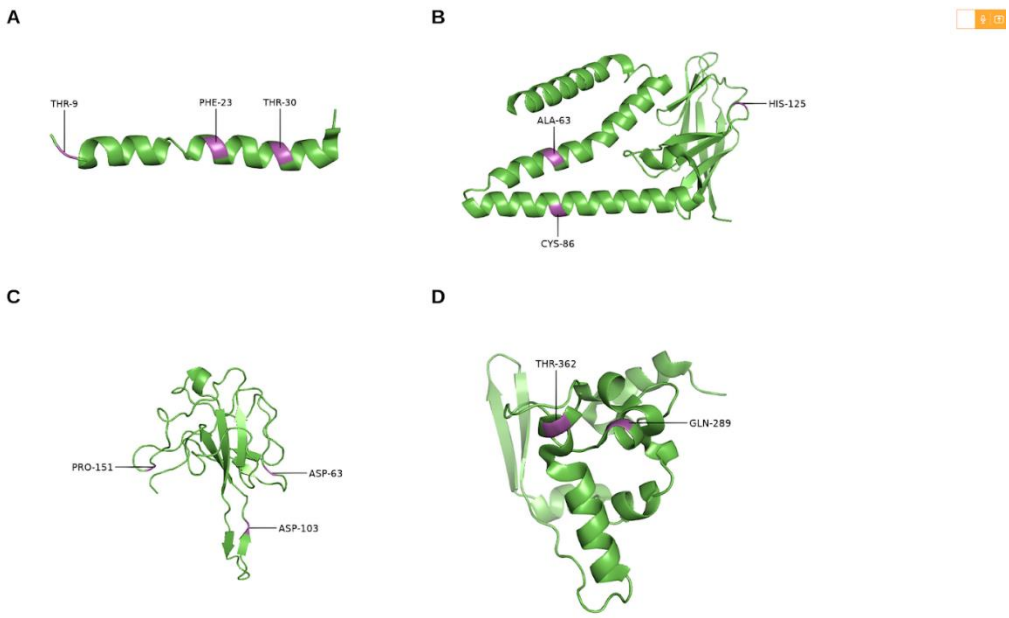


Figure 6. Positively selected sites in E, M and N protein structures. (A) Transmembrane domain from Envelope protein (7K3G chain A); (B) Membrane protein (8CTK, chain A); (C) N-terminal domain from Nucleocapsid protein (7VNU, chain A); (D) C-terminal domain from Nucleocapsid protein (6ZCO, chain A).

In the transmembrane domain of protein E, all evaluated amino acid mutations lead to the increase of the molecular flexibility, despite T9I indicating a stabilizing effect (ΔΔG = 0.321 kcal/mol). For protein M most of the possible alterations promote a stabilizing effect in protein structure, excepting for H125Q (ΔΔG = -0.054 kcal/mol). In the N-terminal domain from N protein, most alterations suggest a stabilizing effect and the decrease of the molecular flexibility by reduction of the vibrational energy. Contrary, alterations observed in the C-terminal domain indicate a destabilizing effect with increase of molecular flexibility. Some mutations, such as M:H125L/Q/Y, N:D103N/Y, N:P151L/S, and N:T362I/K show variate stabilizing/destabilizing patterns for the same site.

Table 5. DynaMut estimations for positively selected sites from proteins E, M and N

Protein	Site	Mutation	ΔΔG DynaMut (kcal/mol)	$\Delta\Delta S_{vib}$ ENCoM (kcal.mol^{-1}.K^{-1})
E	9	T → I	0.321 (S)	0.268 (↑MF)
E	23	F → C	-1.047 (D)	0.458 (↑MF)
E	30	T → I	-0.045 (D)	0.180 (↑MF)
M	63	A → T	0.050 (S)	-0.143 (↓MF)
M	63	A → V	0.411 (S)	-0.123 (↓MF)
M	86	C → F	0.059 (S)	-0.098 (↓MF)
M	86	C → S	0.052 (S)	0.020 (↑MF)
M	125	H → L	0.137 (S)	0.498 (↑MF)

M	125	H → Q	-0.054 (D)	0.465 (↑MF)
M	125	H → Y	0.522 (S)	-0.579 (↓MF)
N	63	D → G	0.009 (S)	0.609 (↑MF)
N	63	D → Y	2.025 (S)	-0.582 (↓MF)
N	103	D → N	0.210 (S)	-0.063 (↓MF)
N	103	D → Y	-0.014 (D)	-0.076 (↓MF)
N	151	P → L	0.776 (S)	-0.110 (↓MF)
N	151	P → S	-0.623 (D)	0.713 (↑MF)
N	289	Q → L	-0.035 (D)	0.127 (↑MF)
N	289	Q → H	-0.392 (D)	0.292 (↑MF)
N	362	T → I	-0.170 (D)	0.095 (↑MF)
N	362	T → K	1.056 (S)	-0.208 (↓MF)

Discussion

Rio Grande do Sul is currently the fourth state most affected by COVID-19 in Brazil (https://covid.saude.gov.br/ access in July 26). The P.1 lineage initiated a new wave of infections in Brazil around November 2020, starting in Manaus (northern Brazil) and spreading across the country (Faria et al., 2021). In RS, P.1 arrived in mid of January 2021, keeping a massive transmission until April 2021, characterizing the third COVID-19 wave in the state (Varela et al., 2021). According to the phylogenomic analysis, the genomes from RS state formed monophyletic groups for most of the lineages, with specific clades to lineages and sublineages belonging to Alpha, Gamma, Delta and Omicron VOCs. This data may suggest some intra-lineage genetic conservation in the SARS-CoV-2 genomes from RS state. However, as seen in AudacityInstant data search, four of our sequenced samples are most closely related to genomes from Chile, Mexico, Canada and USA. This data can suggest viral transmission events introducing these different SARS-

CoV-2 samples to RS or exporting them to other regions. In fact, except for genome RS-44479 which was collected in June, 2021 and had their closest related sequence dated to February, 2021, in Rio de Janeiro, all remaining sequences have older collection dates than their matches, potentially suggesting viral exportation events from RS state.

RNA viruses have a higher mutation rate than DNA viruses and organisms (Duffy, 2018). Selective pressure occurs in a way that the virus can keep its transmission and immune evasion mechanisms updated according to the host characteristics (Zarai et al., 2020). The E protein, presents the smallest structural protein of SARS-CoV-2 and keeps their structure highly conserved across diverse genres of β-coronaviruses (Yadav et al., 2021). The E protein comprises three main domains, the N-terminal (NT), C-terminal (CT) and transmembrane domain (TMD). As observed by Emam and colleagues (2021), Envelope protein sites 66 and 68, located inside a putative transmembrane α-helical domain C-terminal part are found under adaptive selective pressure. The amino acid substitution of a threonine by isoleucine in site 9, located in TMD, and which was found to be positively selected in this study, could indicate an increased interaction with membrane lipids, as well as, the alteration of the capacity of membrane attachment and ER targeting by the E protein (Timmers et al., 2021). Sites 72 and 73, in turn, located at D-L-L-V motif, bind to the host protein PALS1, facilitating infection (Timmers et al., 2021). In this way, the impact of new mutations at these positively selected sites remains to be elucidated.

The M protein has a great importance in the mounting of the virion and the other structural proteins in the coronaviruses (Neuman et al., 2011). In SARS-CoV-

2, that protein can be related in antigenic reactions, with the S and N proteins (Lopandić et al., 2021) even reducing the interferon I responses (Sui et al., 2021), being an important piece in the viral genome. Therefore, modifications in his genome structure can be crucial to the virus survival, being probably the reason for the low incidence of adaptive selection in that protein.

The N protein structure is composed of three main domains: N-terminal domain (NTD), a linker domain rich on serine and arginine residues (SR-rich linker), and a C-terminal domain (CTD) (Timmers et al., 2021). NTD and CTD comprise major antigenic sites of the N protein in SARS-CoV virus (Surjit & Lal, 2009). This protein has with function the packing of viral genetic material besides works in the immune escape, blocking interferons and other defense mechanisms of the host (Bai et al., 2021). According to Rahman et al., the large number of alterations in that protein makes it difficult to create vaccines and medications that could use it as a target (Rahman et al., 2020). The co-occurring amino acid mutations R203K and G204R are known to enhance replication, fitness, and pathogenesis of SARS-CoV-2 (Johnson et al., 2022). However, only site 204 presents evidence to be under adaptive selection in this study.

Changes in nucleotides can modify the structure of the protein, increasing or decreasing their stability (Jaenicke, 1996) and a stable protein is “stronger”, because the stability difficult their denaturation under high temperatures or other inhospitable situations (Sikosek & Chan, 2014). Whereas, the flexibility of a protein is related to its functions and conformation (Zhao, 2010). According to the DynaMut results about the E protein, although both analyzed mutations increase the molecular flexibility, a destabilizing effect was related to F23C and T30I mutations.

These mutations have low frequency (≅0.001%) on SARS-CoV-2 genomes from GISAID. Mutation F23C can be found more frequently in Omicron lineage (67.12% of the F23C occurrences in the world), but also described on Delta, Alpha and Gamma lineages, while T30I is present in these same VOCs (≅0.010%), being mainly associated with Delta lineage (20.28% of the T30I occurrences in the world).

In M protein, we analyzed three sites with different mutations among them. Sites 63 and 86 achieved a stabilizing effect in all tested polymorphisms, whereas in site 125, one of the three analyzed mutations demonstrated destabilizing impact. The mutation H125Y is the most frequent on GISAID with 13,660 occurrences in SARS-CoV-2 genomes in the world. Present in all variants of concern, this mutations was found to be prevalently associated to the Delta lineage (36.55% of H125Y occurrences), while the variant H125L is less spread, occurring in ≅0.001% of world genomes, specially in Delta lineage. Another minor variant, H125Q (≅0.0009%), also leads to a destabilizing effect and is mostly related to Delta and other VOCs such as Alpha, Gamma and Omicron. Together with C86S, minor variants H125L and H125Q seem to increase the protein flexibility.

The N-terminal domain of N protein had 3 sites analyzed with 2 different mutations each. Most mutations lead to a stabilizing effect on the protein, except for mutations D103Y and P151S. Interestingly, mutation D103Y (≅0.039%), which alters the negatively charged aspartic acid by the polar amino acid tyrosine is three times more frequent than mutation D103N (≅0.010%), which brought a stabilizing effect to this protein domain structure alteration to the polar amino acid asparagine. While D103Y is more frequently found in Delta lineage (20.5% of D103Y occurrences), D103N is more associated with Omicron (41.33% of D103N

occurrences), among all VOCs. However, both mutations favor molecular rigidification. P151S (≅0.77% of SARS-CoV-2 genomes in the world), in turn, increases protein flexibility by alteration of a non-polar proline by a polar serine. Prevalent in Omicron (92.35% of P151S occurrences), P151S is more frequent than the P151L (non-polar proline to non-polar leucine), which stabilizes protein structure, with more than 25,000 occurrences on SARS-CoV-2 genomes in the world.

In the C-terminal domain of N protein are also considered 2 sites with two different mutations each, and the majority of them seem to destabilize the protein. Only mutation T362K (≅0.035) demonstrates a stabilizing effect on the C-terminal domain structure. Surprisingly, as observed in N-terminal domain mutations, the polymorphism T362I (≅0.25%) that lead to a destabilizing effect is seven times more frequent than the substitution for a lysine residue. Despite destabilizing the protein, T362K is also related (as T362I) to the increase of protein flexibility, being the molecular flexibility of N protein in SARS-CoV-2 directly connected to the affinity with medications (Matsuo, 2021).

Similar results were found by Rahman and colleagues (2020) in the analysis of the structural effects of mutations D103Y and Q289H. For T362I, their results indicate a stabilizing effect, contradicting our findings. However, both agree that this mutation is associated with the increase of molecular flexibility.

By last, despite our findings, more studies are necessary to understand the genomic and structural changes that favor the adaptation of SARS-CoV-2 virus in host-pathogen interaction, as well as, study viral patterns of SARS-CoV-2 in Rio Grande do Sul state.

Declarations

Availability of data and materials

Full tables acknowledging the authors and corresponding labs submitting sequencing data used in this study can be found in Supplementary File 3. Additional information used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare no competing interests.

Funding

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)* - Finance code 001

Author's contributions

Conceptualization: Patrícia Aline Gröhs Ferrareze, Claudia Elizabeth Thompson; **Methodology:** Amanda de Menezes Mayer, Patrícia Aline Gröhs Ferrareze, Luiz Felipe Valter de Oliveira, Tatiana Schäffer Gregianini, Claudia Elizabeth Thompson**;** **Formal analysis and investigation:** Amanda de Menezes Mayer, Patrícia Aline Gröhs Ferrareze, Gabriel Dickin Caldana, Claudia Elizabeth Thompson**;** **Availability of Samples:** Tatiana Schäffer Gregianini; **Writing - original draft preparation:** Amanda de Menezes Mayer, Patrícia Aline Gröhs Ferrareze, Gabriel Dickin Caldana, Claudia Elizabeth Thompson**;** **Writing - review and editing:**

Amanda de Menezes Mayer, Patrícia Aline Gröhs Ferrareze, Gabriel Dickin Caldana, Luiz Felipe Valter de Oliveira, Tatiana Schäffer Gregianini, Lívia Kmetzsch, Claudia Elizabeth Thompson**; Funding acquisition:** Claudia Elizabeth Thompson, Luiz Felipe Valter de Oliveira**; Resources:** Claudia Elizabeth Thompson, Tatiana Schäffer Gregianini, Luiz Felipe Valter de Oliveira**; Supervision:** Claudia Elizabeth Thompson**.** All authors have read and approved the manuscript.

Acknowledgements

We thank the administrators of the GISAID database and research groups across the world for supporting the rapid and transparent sharing of genomic data during the COVID-19 pandemic and the *Governo do Estado do Rio Grande do Sul* and *Ministério da Saúde* for supplies and equipment used in the SARS-CoV-2 diagnosis routine.

References

Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*, *60*(5), 685–699. https://doi.org/10.1093/sysbio/syr041

Bai, Z., Cao, Y., Liu, W., & Li, J. (2021). The sars-cov-2 nucleocapsid protein and its role in viral structure, biological functions, and a potential target for drug or vaccine mitigation. *Viruses*, *13*(6). https://doi.org/10.3390/v13061115

Chakraborty, S. (2022). E484K and N501Y SARS-CoV 2 spike mutants Increase ACE2 recognition but reduce affinity for neutralizing antibody. *International Immunopharmacology*, *102*, 108424. https://doi.org/10.1016/j.intimp.2021.108424

Chen, J., Wang, R., Gilby, N. B., & Wei, G.-W. (2022). Omicron variant (B.1.1.529): Infectivity, vaccine breakthrough, and antibody resistance. *Journal of Chemical Information and Modeling*, *62*(2), 412–422. https://doi.org/10.1021/acs.jcim.1c01451

Chen, R. E., Zhang, X., Case, J. B., Winkler, E. S., Liu, Y., VanBlargan, L. A., Liu, J., Errico, J. M., Xie, X., Suryadevara, N., Gilchuk, P., Zost, S. J., Tahan, S., Droit, L., Turner, J. S., Kim, W., Schmitz, A. J., Thapa, M., Wang, D., … Diamond, M. S. (2021). Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nature Medicine*, *27*(4), 717–726. https://doi.org/10.1038/s41591-021-01294-w

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*(5), 533–534. https://doi.org/10.1016/s1473-3099(20)30120-1

Duffy, S. (2018). Why are RNA virus mutation rates so damn high? *PLoS Biology*, *16*(8). https://doi.org/10.1371/journal.pbio.3000003

Eden, J.-S., & Sim, E. (2020). *SARS-CoV-2 Genome Sequencing Using Long Pooled Amplicons on Illumina Platforms v1*. ZappyLab, Inc. http://dx.doi.org/10.17504/protocols.io.befyjbpw

Elbe, S., & Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*, *1*(1), 33–46. https://doi.org/10.1002/gch2.1018

Emam, M., Oweda, M., Antunes, A., & El-Hadidi, M. (2021). Positive selection as a key player for SARS-CoV-2 pathogenicity: Insights into ORF1ab, S and E genes. *Virus Research*, *302*, 198472. https://doi.org/10.1016/j.virusres.2021.198472

Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., Candido, D. da S., Mishra, S., Crispim, M. A. E., Sales, F. C. S., Hawryluk, I., McCrone, J. T., Hulswit, R. J. G., Franco, L. A. M., Ramundo, M. S., de Jesus, J. G., Andrade, P. S., Coletti, T. M., Ferreira, G. M., Silva, C. A. M., Manuli, E. R., … Sabino, E. C. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science (New York, N.y.)*, *372*(6544). https://doi.org/10.1126/science.abh2644

Franceschi, V. B., Caldana, G. D., Perin, C., Horn, A., Peter, C., Cybis, G. B., Ferrareze, P. A. G., Rotta, L. N., Cadegiani, F. A., Zimerman, R. A., & Thompson, C. E. (2021). Predominance of the sars-cov-2 lineage P.1 And its sublineage P.1.2 in patients from the metropolitan region of Porto

Alegre, southern Brazil in march 2021. *Pathogens*, *10*(8), 988. https://doi.org/10.3390/pathogens10080988

Ghosh, S., & Chakraborty, S. (2020). Phylogenomics analysis of sars-cov2 genomes reveals distinct selection pressure on different viral strains. *BioMed Research International*, *2020*. https://doi.org/10.1155/2020/5746461

Giovanetti, M., Fonseca, V., Wilkinson, E., Tegally, H., San, E. J., Althaus, C. L., Xavier, J., Nanev Slavov, S., Viala, V. L., Ranieri Jerônimo Lima, A., Ribeiro, G., Souza-Neto, J. A., Fukumasu, H., Lehmann Coutinho, L., Venancio da Cunha, R., Freitas, C., Campelo de A e Melo, C. F., Navegantes de Araújo, W., Do Carmo Said, R. F., … de Alcantara, L. C. J. (2022). Replacement of the Gamma by the Delta variant in Brazil: Impact of lineage displacement on the ongoing pandemic. *Virus Evolution*, *8*(1). https://doi.org/10.1093/ve/veac024

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of phyml 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Gularte, J. S., da Silva, M. S., Mosena, A. C. S., Demoliner, M., Hansen, A. W., Filippi, M., Pereira, V. M. de A. G., Heldt, F. H., Weber, M. N., de Almeida, P. R., Hoffmann, A. T., Valim, A. R. de M., Possuelo, L. G., Fleck, J. D., & Spilki, F. R. (2022). Early introduction, dispersal and evolution of Delta SARS-CoV-2 in Southern Brazil, late predominance of AY.99.2 and AY.101 related lineages. *Virus Research*, *311*, 198702. https://doi.org/10.1016/j.virusres.2022.198702

Jaenicke, R. (1996). How do proteins acquire their three-dimensional structure and stability? *Naturwissenschaften*, *83*(12), 544–554. https://doi.org/10.1007/bf01141979

Johnson, B. A., Zhou, Y., Lokugamage, K. G., Vu, M. N., Bopp, N., Crocquet-Valdes, P. A., Kalveram, B., Schindewolf, C., Liu, Y., Scharton, D., Plante, J. A., Xie, X., Aguilar, P., Weaver, S. C., Shi, P.-Y., Walker, D. H., Routh, A. L., Plante, K. S., & Menachery, V. D. (2022). Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. *PLOS Pathogens*, *18*(6), e1010627. https://doi.org/10.1371/journal.ppat.1010627

Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*(5), 1208–1222. https://doi.org/10.1093/molbev/msi105

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lopandić, Z., Protić-Rosić, I., Todorović, A., Glamočlija, S., Gnjatović, M., Ćujic, D., & Gavrović-Jankulović, M. (2021). IgM and igg immunoreactivity of sars-cov-2 recombinant M protein. *International Journal of Molecular Sciences*, *22*(9). https://doi.org/10.3390/ijms22094951

Matsuo, T. (2021). Viewing sars-cov-2 nucleocapsid protein in terms of molecular flexibility. *Biology*, *10*(6). https://doi.org/10.3390/biology10060454

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., Droese, B., Klaus, J. P., Makino, S., Sawicki, S. G., Siddell, S. G., Stamou, D. G., Wilson, I. A., Kuhn, P., & Buchmeier, M. J. (2011). A structural analysis of M protein in coronavirus assembly and morphology. *Journal of Structural Biology*, *174*(1). https://doi.org/10.1016/j.jsb.2010.11.021

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Okonechnikov, K., Golosova, O., & Fursov, M. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Pond, S. L. K., Frost, S. D. W., & Muse, S. V. (2004). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*, *21*(5), 676–679. https://doi.org/10.1093/bioinformatics/bti079

Rahman, M. S., Islam, M. R., Alam, A. S. M. R. U., Islam, I., Hoque, M. N., Akter, S., Rahaman, Md. M., Sultana, M., & Hossain, M. A. (2020). Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *Journal of Medical Virology*, *93*(4), 2177–2195. https://doi.org/10.1002/jmv.26626

Rodrigues, C. H., Pires, D. E., & Ascher, D. B. (2018). DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research*, *46*(W1), W350–W355. https://doi.org/10.1093/nar/gky300

Satarker, S., & Nampoothiri, M. (2020). Structural proteins in severe acute respiratory syndrome coronavirus-2. *Archives of Medical Research*, *51*(6), 482–491. https://doi.org/10.1016/j.arcmed.2020.05.012

Shiehzadegan, S., Alaghemand, N., Fox, M., & Venketaraman, V. (2021). Analysis of the delta variant B.1.617.2 COVID-19. *Clinics and Practice*, *11*(4). https://doi.org/10.3390/clinpract11040093

Sikosek, T., & Chan, H. S. (2014). Biophysics of protein evolution and evolutionary protein biophysics. *Journal of The Royal Society Interface*, *11*(100), 20140419. https://doi.org/10.1098/rsif.2014.0419

Sui, L., Zhao, Y., Wang, W., Wu, P., Wang, Z., Yu, Y., Hou, Z., Tan, G., & Liu, Q. (2021). SARS-CoV-2 membrane protein inhibits type I interferon production through ubiquitin-mediated degradation of TBK1. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.662989

Surjit, M., & Lal, S. K. (2009). The nucleocapsid protein of the SARS coronavirus: Structure, function and therapeutic potential. In *Molecular Biology of the SARS-Coronavirus* (pp. 129–151). Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-03683-5_9

Timmers, L. F. S. M., Peixoto, J. V., Ducati, R. G., Bachega, J. F. R., de Mattos Pereira, L., Caceres, R. A., Majolo, F., da Silva, G. L., Anton, D. B., Dellagostin, O. A., Henriques, J. A. P., Xavier, L. L., Goettert, M. I., & Laufer, S. (2021). SARS-CoV-2 mutations in Brazil: From genomics to putative clinical conditions. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-91585-6

Varela, A. P. M., Prichula, J., Mayer, F. Q., Salvato, R. S., Sant'Anna, F. H., Gregianini, T. S., Martins, L. G., Seixas, A., & Veiga, A. B. G. da. (2021). SARS-CoV-2 introduction and lineage dynamics

across three epidemic peaks in Southern Brazil: Massive spread of P.1. *Infection, Genetics and Evolution*, *96*. https://doi.org/10.1016/j.meegid.2021.105144

Wang, X., & Powell, C. A. (2021). How to translate the knowledge of COVID-19 into the prevention of Omicron variants. *Clinical and Translational Medicine*, *11*(12). https://doi.org/10.1002/ctm2.680

Wink, P. L., Ramalho, R., Monteiro, F. L., Volpato, F. C. Z., Willig, J. B., Lovison, O. von A., Zavascki, A. P., Barth, A. L., & Martins, A. F. (2022). Genomic surveillance of sars-cov-2 lineages indicates early circulation of P.1 (Gamma) variant of concern in southern Brazil. *Microbiology Spectrum*, *10*(1). https://doi.org/10.1128/spectrum.01511-21

Yadav, R., Chaudhary, J. K., Jain, N., Chaudhary, P. K., Khanra, S., Dhamija, P., Sharma, A., Kumar, A., & Handu, S. (2021). Role of structural and non-structural proteins and therapeutic targets of sars-cov-2 for COVID-19. *Cells*, *10*(4), 821. https://doi.org/10.3390/cells10040821

Zarai, Y., Zafrir, Z., Siridechadilok, B., Suphatrakul, A., Roopin, M., Julander, J., & Tuller, T. (2020). Evolutionary selection against short nucleotide sequences in viruses and their related hosts. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *27*(2). https://doi.org/10.1093/dnares/dsaa008

Zhao, Q. (2010). Protein flexibility as a biosignal. *Critical ReviewsTM in Eukaryotic Gene Expression*, *20*(2), 157–170. https://doi.org/10.1615/critreveukargeneexpr.v20.i2.60

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, *382*(8), 727–733. https://doi.org/10.1056/nejmoa2001017

3. CONCLUSÕES

- Nossos estudos trazem importantes descrições sobre a dispersão do SARS-CoV-2 no município de Esteio, bem como descrevem a evolução genômica do vírus.
- Além de outras mutações, descrevemos a primeira aparição, até o momento da redação deste manuscrito, da mutação E484K em amostras da linhagem P.1 dentro do estado. Logo em seguida, esta mutação se mostrou extremamente importante por estar relacionada ao escape imunológico. Em virtude de eventos evolutivos independentes e consequente vantagem evolutiva, E484K se manteve em diferentes linhagens virais e se espalhou rapidamente pelo país.
- Considerando nossas análises, as linhagens Gamma e Delta foram as mais prevalentes no estado durante os dois anos iniciais da pandemia e nossos genomas virais demonstraram uma conservação genética intra-linhagem, junto de diversos grupos monofiléticos, apesar da provável importação viral devido a sua relação com genomas de diversas localidades.
- As sequências genômicas de amostras virais coletadas no RS tiveram diversos sítios sob seleção adaptativa em suas proteínas estruturais, levando aumento de diversidade nos sítios submetidos à seleção e dificultando a manutenção da eficácia dos imunizantes. Ademais, diferentes mutações indicaram desestabilização da estrutura viral em variados sítios das proteínas analisadas, muitas vezes sendo estas mutações mais frequentes do que os sítios estabilizados.

4. PERSPECTIVAS

- Em virtude do surgimento de novas linhagens, torna-se importante manter as análises genômicas do SARS-CoV-2 em atualização, para que haja o entendimento das mutações e modificações na estrutura viral, bem como para auxiliar no desenvolvimento de novos tratamentos e vacinas contra o SARS-CoV-2.
- Com estes estudos, foi possível compreender um pouco mais sobre algumas particularidades genômicas e dispersão do vírus no estado. No entanto, estudos adicionais são necessários para entendimento da evolução viral do SARS-CoV-2 no Rio Grande do Sul.

REFERÊNCIAS BIBLIOGRÁFICAS

BERRIO, A.; GARTNER, V.; WRAY, G. A. Positive selection within the genomes of SARS-CoV-2 and other Coronaviruses independent of impact on protein function. **PeerJ**, v. 8, 16 out. 2020.

CHEN, M. et al. Safety of SARS-CoV-2 vaccines: a systematic review and meta-analysis of randomized controlled trials. **Infectious Diseases of Poverty**, v. 10, n. 1, 1 dez. 2021.

CUI, J.; LI, F.; SHI, Z. L. Origin and evolution of pathogenic coronaviruses. **Nature reviews. Microbiology**, v. 17, n. 3, p. 181–192, 1 mar. 2019.

DHAR, M. S. et al. Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. **Science**, v. 374, n. 6570, p. 995–999, 19 nov. 2021.

FARIA, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. **Science**, v. 372, n. 6544, 21 maio 2021.

FARIA, N. R. et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. **Virological**, 13 jan. 2021. Disponível em: <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>. Acesso em: 18 ago. 2022.

GANGAVARAPU, K. et al. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. **medRxiv**, p. 2022.01.27.22269965, 2 jun. 2022.

GREANEY, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. **Cell Host & Microbe**, v. 29, n. 3, p. 463, 3 mar. 2021.

HE, X. et al. SARS-CoV-2 Omicron variant: Characteristics and prevention. **MedComm**, v. 2, n. 4, p. 838, 1 dez. 2021.

HU, B. et al. Bat origin of human coronaviruses. **Virology Journal**, v. 12, n. 1, 22 dez. 2015.

IBGE. **Esteio (RS)**. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/rs/esteio.html>. Acesso em: 18 ago. 2022.

KORBER, B. et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. **Cell**, v. 182, n. 4, p. 812, 8 ago. 2020.

KRAMMER, F. SARS-CoV-2 vaccines in development. **Nature 2020 586:7830**, v. 586, n. 7830, p. 516–527, 23 set. 2020.

KSIAZEK, T. G. et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. **New England Journal of Medicine**, v. 348, n. 20, p. 1953–1966, 15 maio 2003.

LAMARCA, A. P. et al. Genomic surveillance of SARS-CoV-2 tracks early interstate transmission of P.1 lineage and diversification within P.2 clade in Brazil. **PLoS Neglected Tropical Diseases**, v. 15, n. 10, 1 out. 2021.

LIU, Y. et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. **Nature**, v. 602, n. 7896, p. 294, 10 fev. 2022.

MARSON, F. A. L.; ORTEGA, M. M. COVID-19 in Brazil. **Pulmonology**, v. 26, n. 4, p. 241, 1 jul. 2020.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **WHO Coronavirus (COVID-19) Dashboard**. Disponível em: <https://covid19.who.int/>. Acesso em: 18 ago. 2022.

NOHYNEK, H.; WILDER-SMITH, A. Does the World Still Need New Covid-19 Vaccines? **The New England Journal of Medicine**, v. 386, n. 22, p. 2140–2142, 2 jun. 2022.

PAG, F. et al. E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. **Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases**, v. 93, 1 set. 2021.

PARK, S. E. Epidemiology, virology, and clinical features of severe acute respiratory syndrome -coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). **Clinical and experimental pediatrics**, v. 63, n. 4, p. 119–124, 2020.

PLANAS, D. et al. Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. **Nature 2021 602:7898**, v. 602, n. 7898, p. 671–675, 23 dez. 2021.

PLANTE, J. A. et al. The variant gambit: COVID-19's next move. **Cell Host and Microbe**, v. 29, n. 4, p. 508–515, 14 abr. 2021.

PREFEITURA DE ESTEIO. **Monitoramento COVID-19 Esteio**. Disponível em: <https://covid.esteio.rs.gov.br/>. Acesso em: 18 ago. 2022.

RAMBAUT, A. et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. **Virological**, 9 dez. 2020. Disponível em: <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>. Acesso em: 18 ago. 2022.

SPGG/RS. **Região Metropolitana de Porto Alegre - RMPA**. Disponível em: <https://atlassocioeconomico.rs.gov.br/regiao-metropolitana-de-porto-alegre-rmpa>. Acesso em: 18 ago. 2022.

SABINO, E. C. et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. **The Lancet**, v. 397, n. 10273, p. 452–455, 6 fev. 2021.

SES/RS. **Paínel Coronavírus RS**. Disponível em: <https://ti.saude.rs.gov.br/covid19/>. Acesso em: 18 ago. 2022.

SILVA, S. J. R. DA; PENA, L. Collapse of the public health system and the emergence of new variants during the second wave of the COVID-19 pandemic in Brazil. **One Health**, v. 13, p. 100287, 1 dez. 2021.

TEGALLY, H. et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. **medRxiv**, p. 2020.12.21.20248640, 22 dez. 2020.

TELENTI, A.; HODCROFT, E. B.; ROBERTSON, D. L. The Evolution and Biology of SARS-CoV-2 Variants. **Cold Spring Harbor Perspectives in Medicine**, v. 12, n. 5, p. a041390, 1 maio 2022.

WANG, R.; CHEN, J.; WEI, G. W. Mechanisms of SARS-CoV-2 Evolution Revealing Vaccine-Resistant Mutations in Europe and America. **Journal of Physical Chemistry Letters**, v. 12, n. 49, p. 11850–11857, 16 dez. 2021.

WOLF, J. M. et al. Temporal spread and evolution of SARS-CoV-2 in the second pandemic wave in Brazil. **Journal of Medical Virology**, v. 94, n. 3, p. 926, 1 mar. 2022.

ZHENG, C. et al. Real-world effectiveness of COVID-19 vaccines: a literature review and meta-analysis. **International Journal of Infectious Diseases**, v. 114, p. 252, 1 jan. 2022.

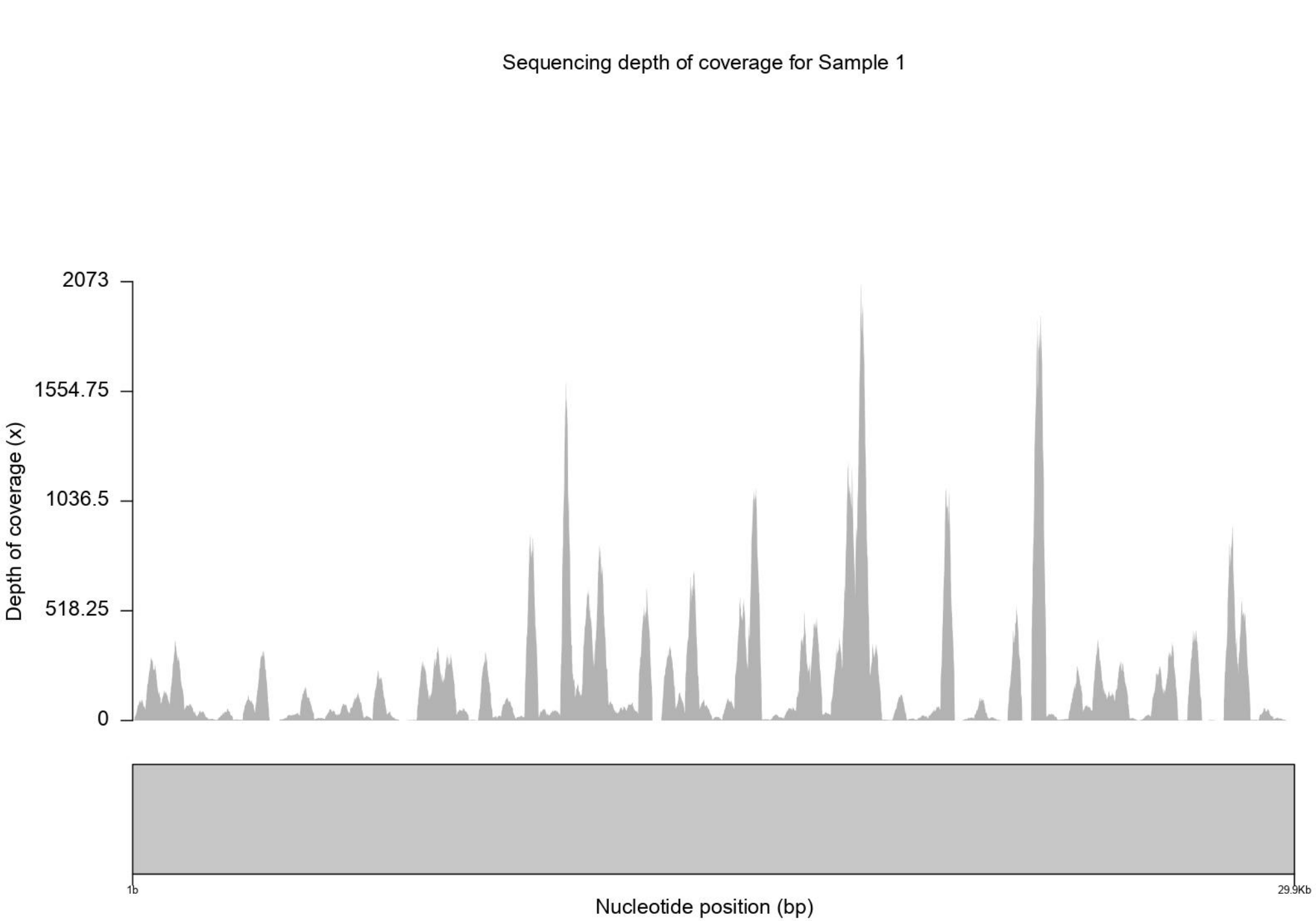
ZHU, N. et al. Brief Report: A Novel Coronavirus from Patients with Pneumonia in China, 2019. **The New England Journal of Medicine**, v. 382, n. 8, p. 727, 2 fev. 2020.

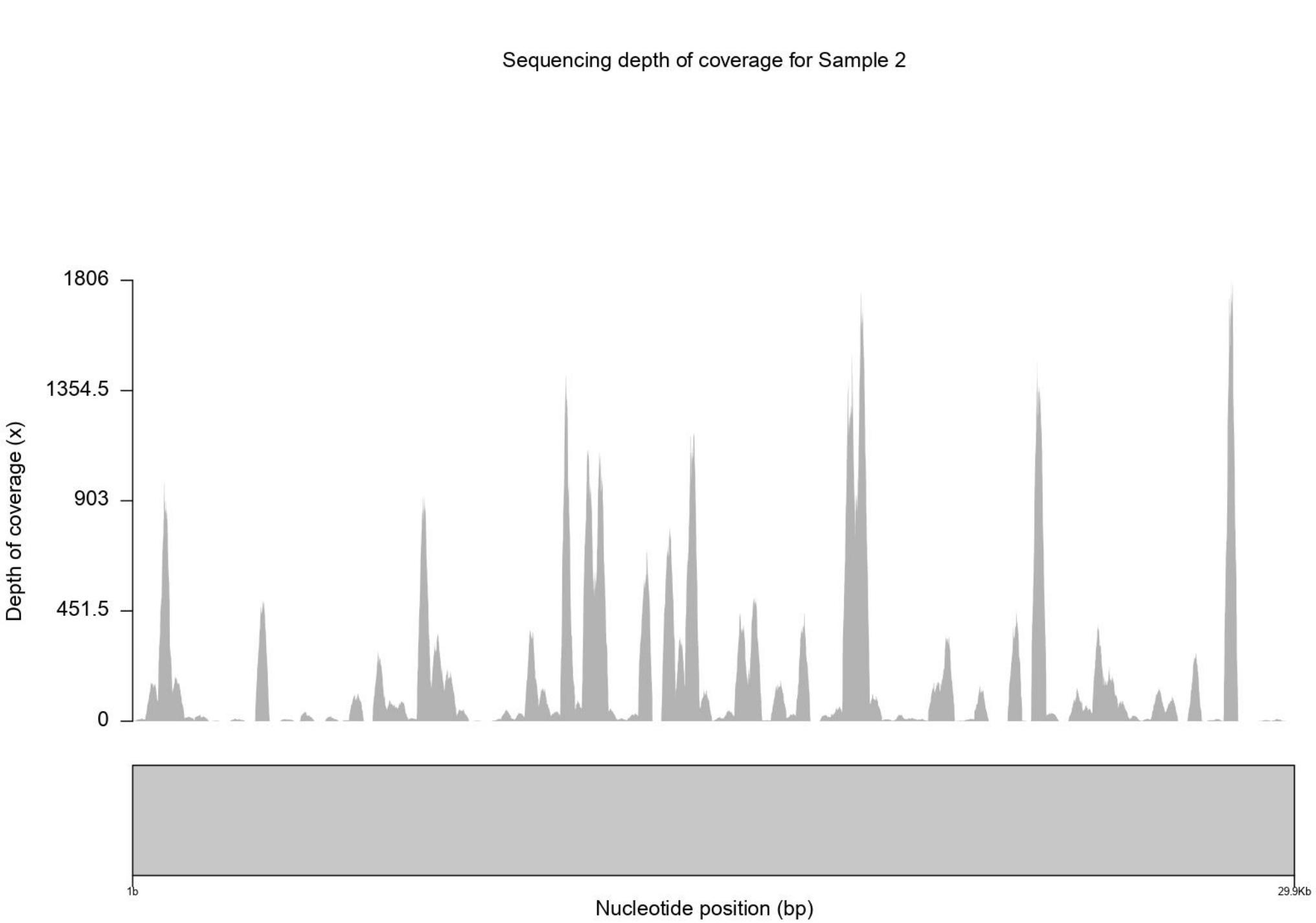
APÊNDICES

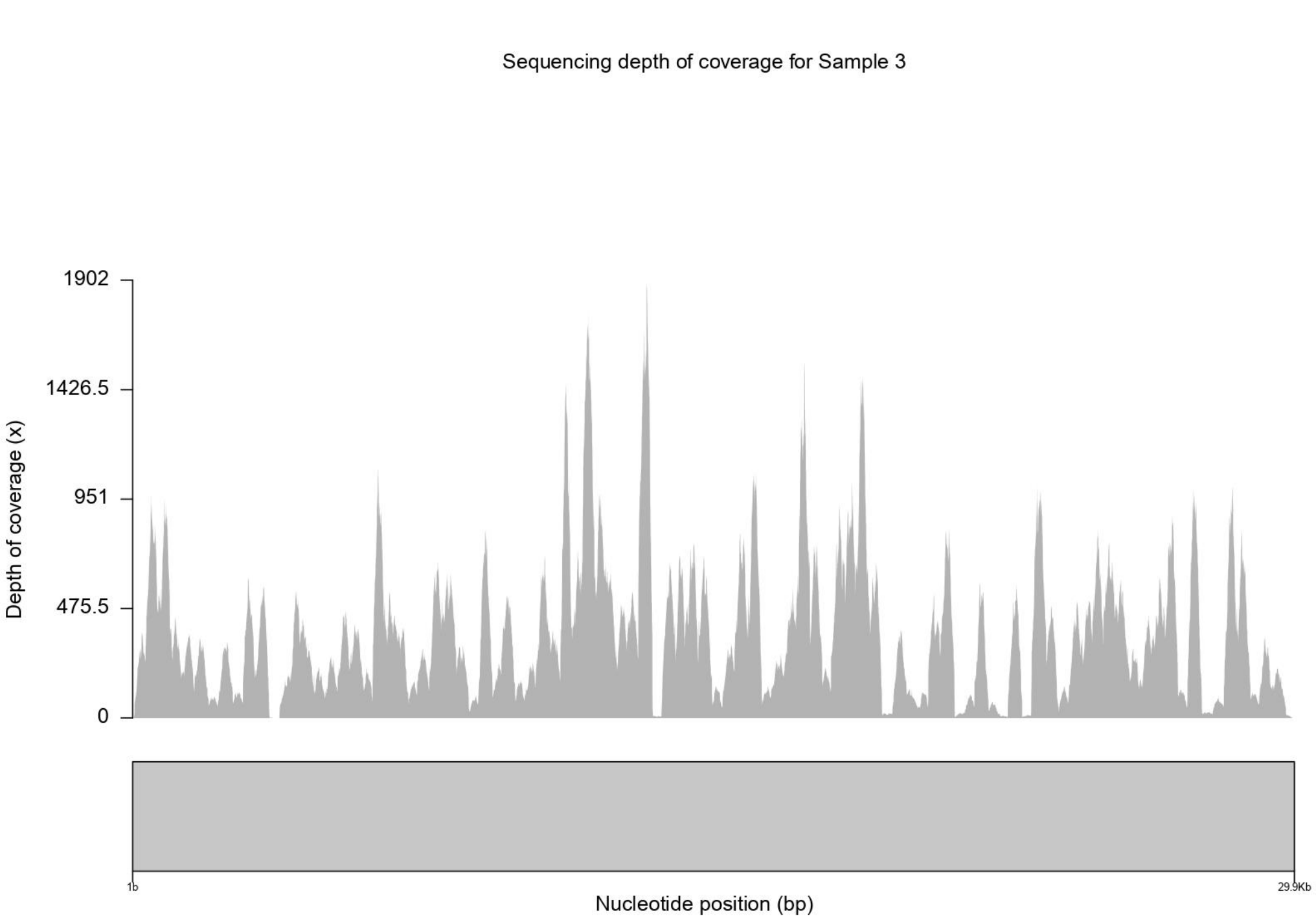
Apresento como Apêndices os materiais suplementares do estudo apresentado no Capítulo I, bem como alguns artigos que foram publicados durante o período do meu mestrado e que são relacionados à COVID-19 e dos quais participei como coautora.

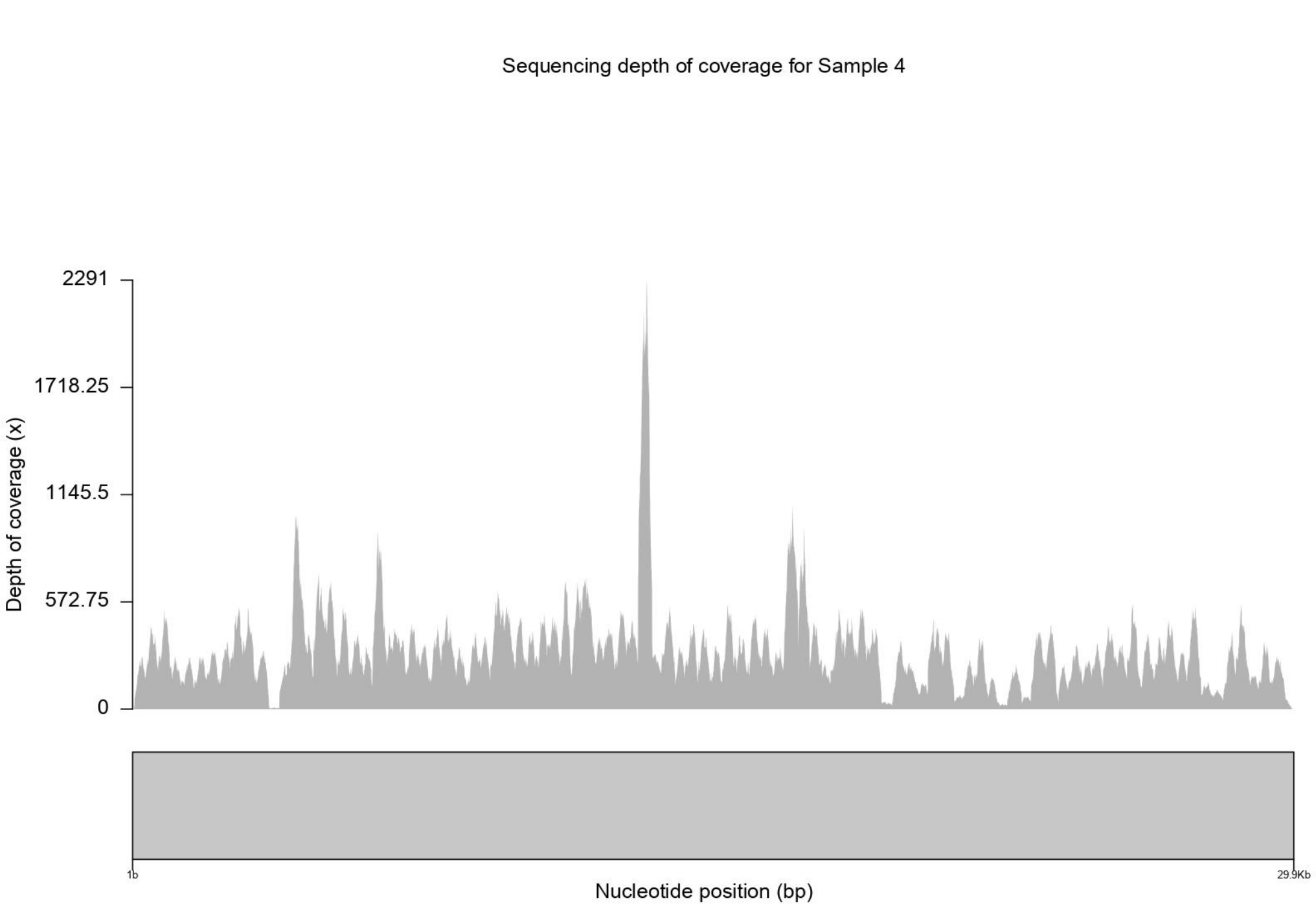
APÊNDICE I

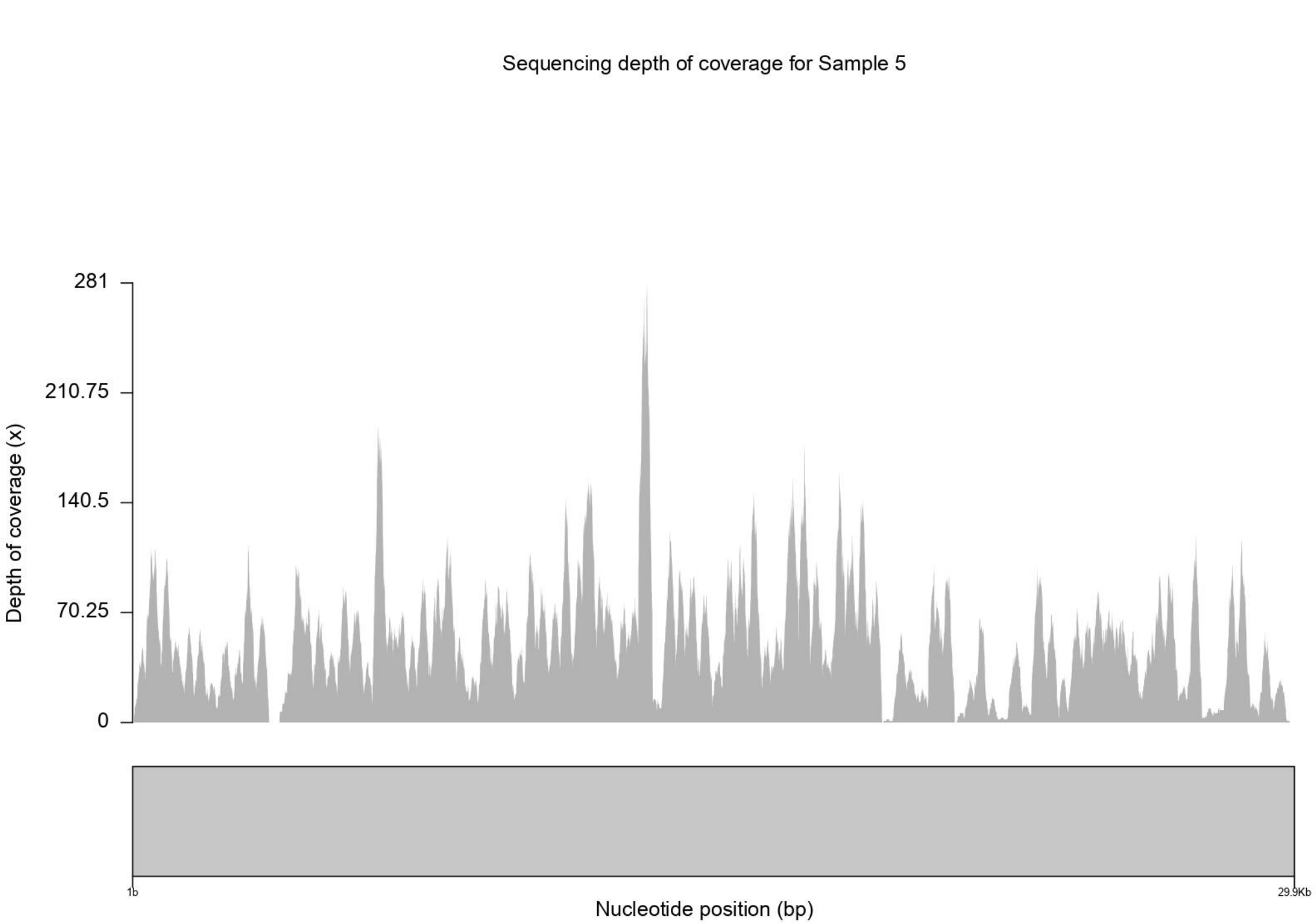
As imagens presentes no Apêndice I constituem o primeiro material suplementar do artigo que compõe o Capítulo I desta dissertação e compreendem os gráficos da cobertura dos doze genomas sequenciados da cidade de Esteio/RS.

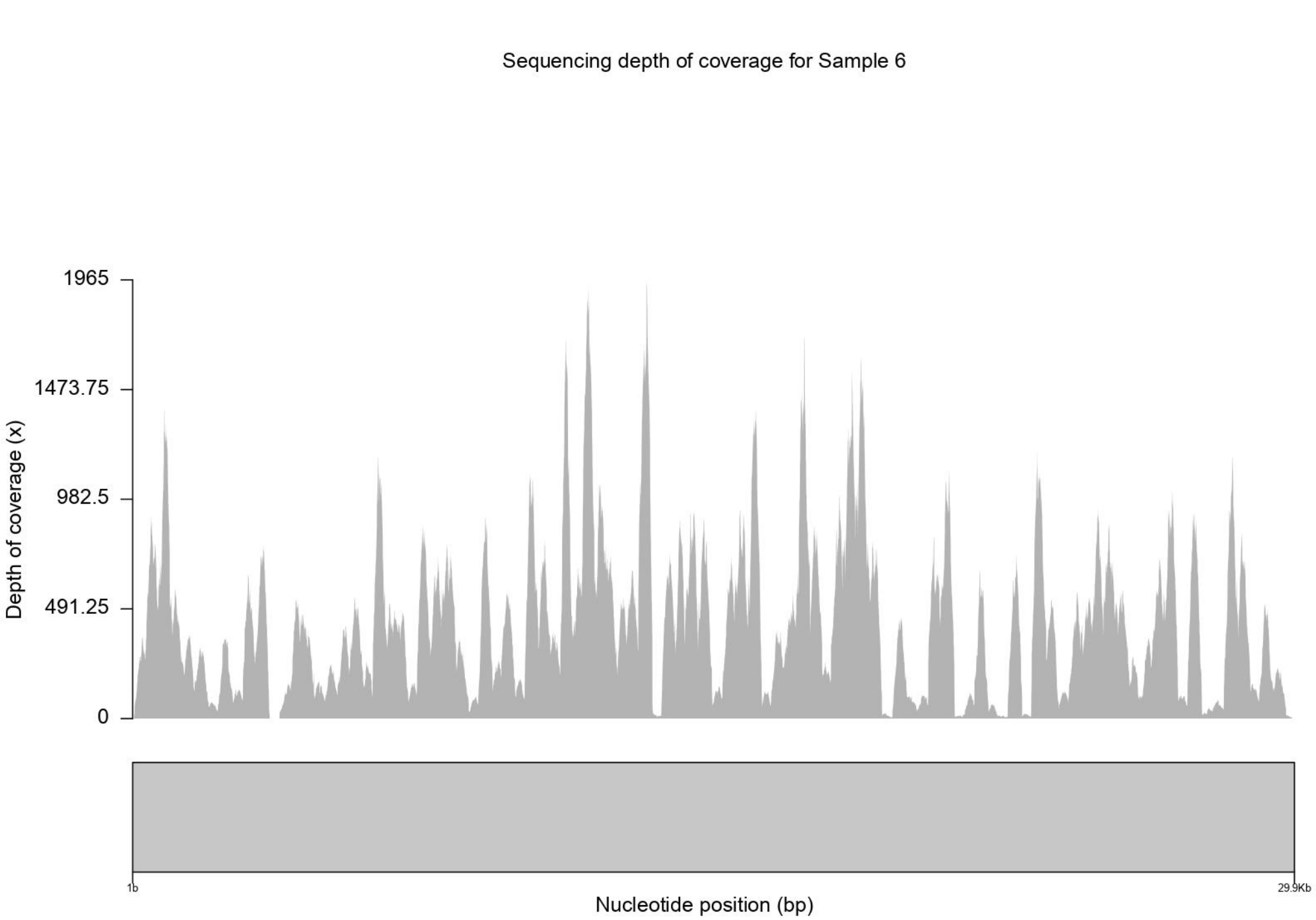

Sequencing depth of coverage for Sample 1
2073
1554.75
1036.5
518.25
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)

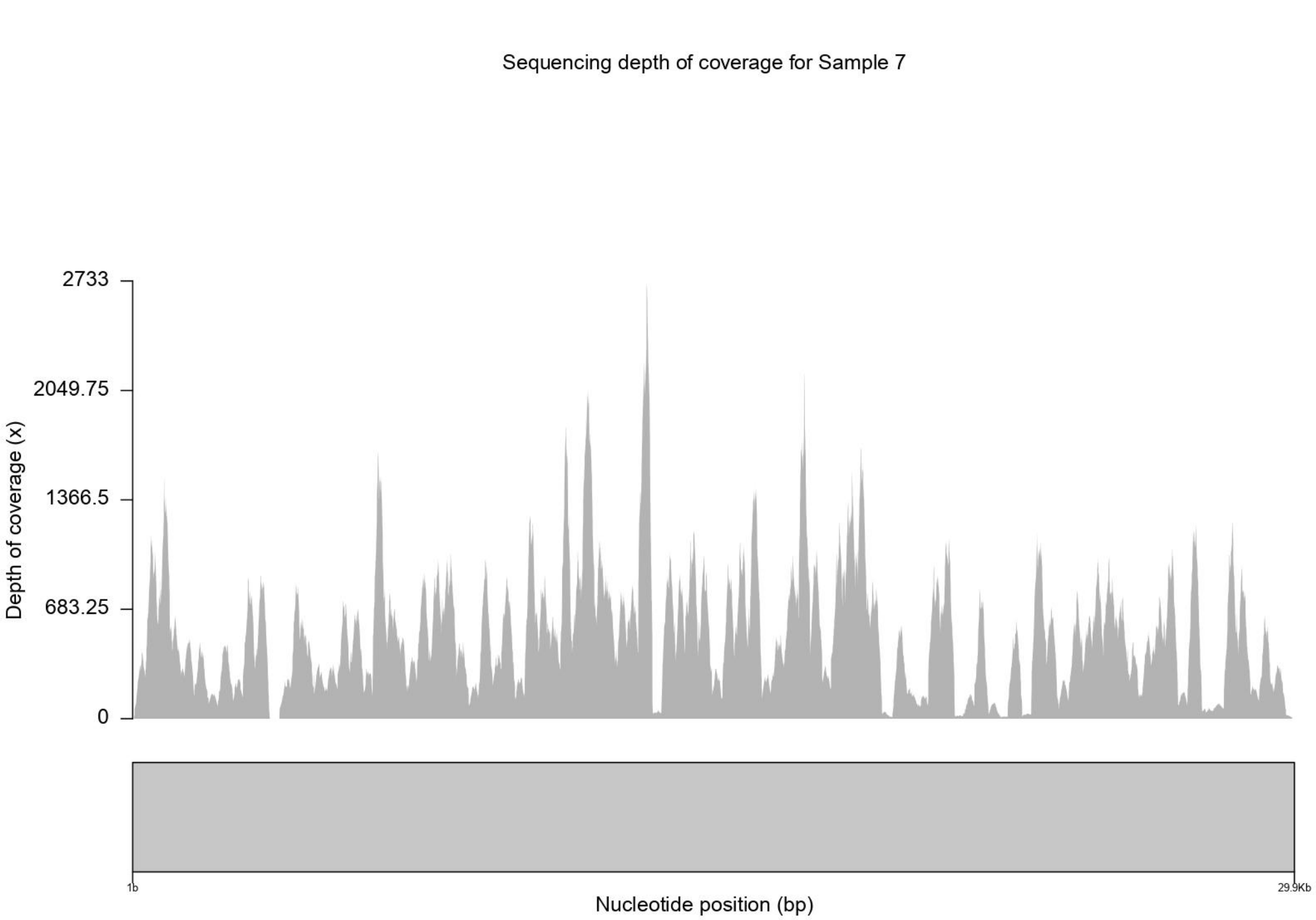

Sequencing depth of coverage for Sample 2
1806
1354.5
903
451.5
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)

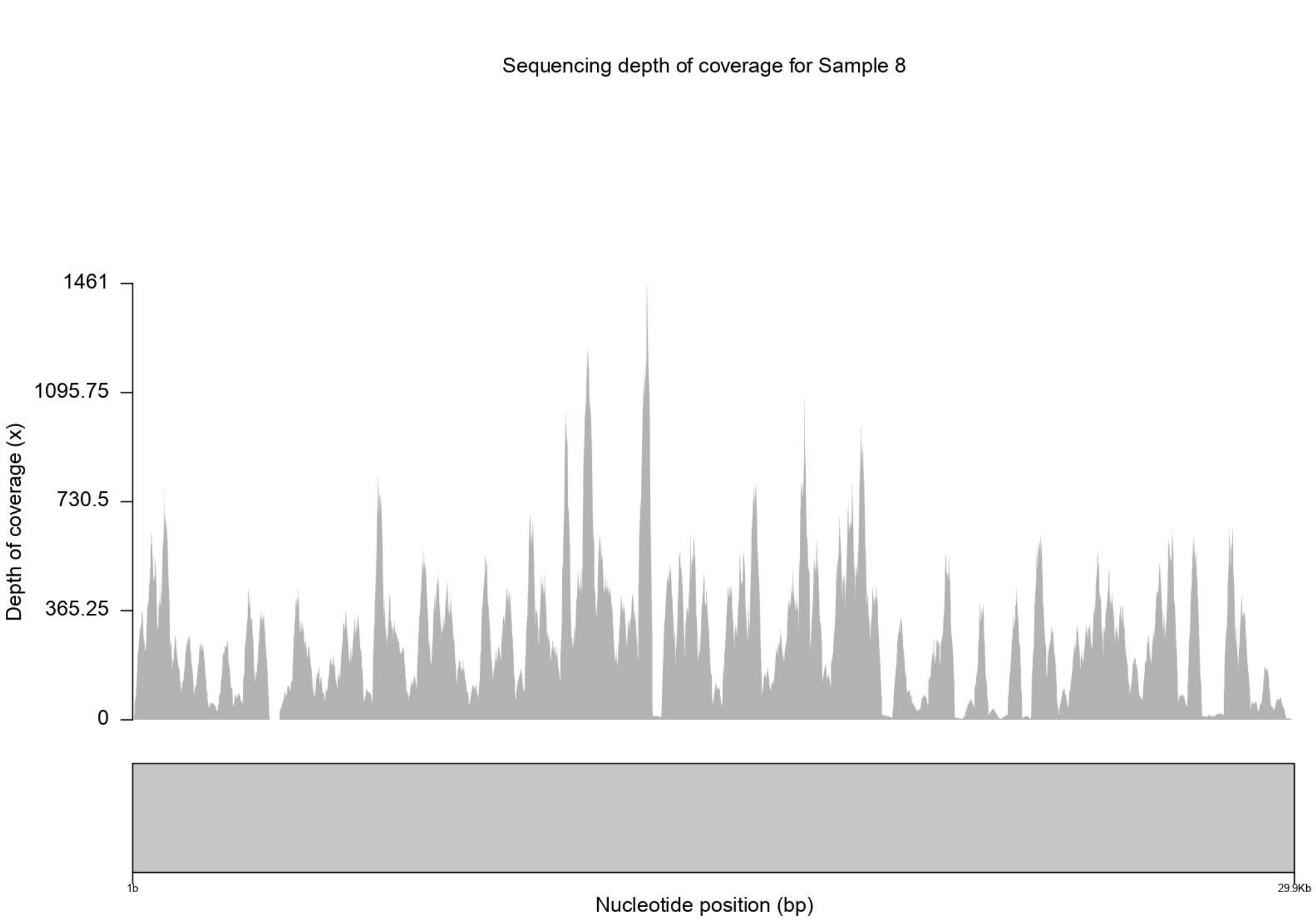

Sequencing depth of coverage for Sample 3
1902
1426.5
951
475.5
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)

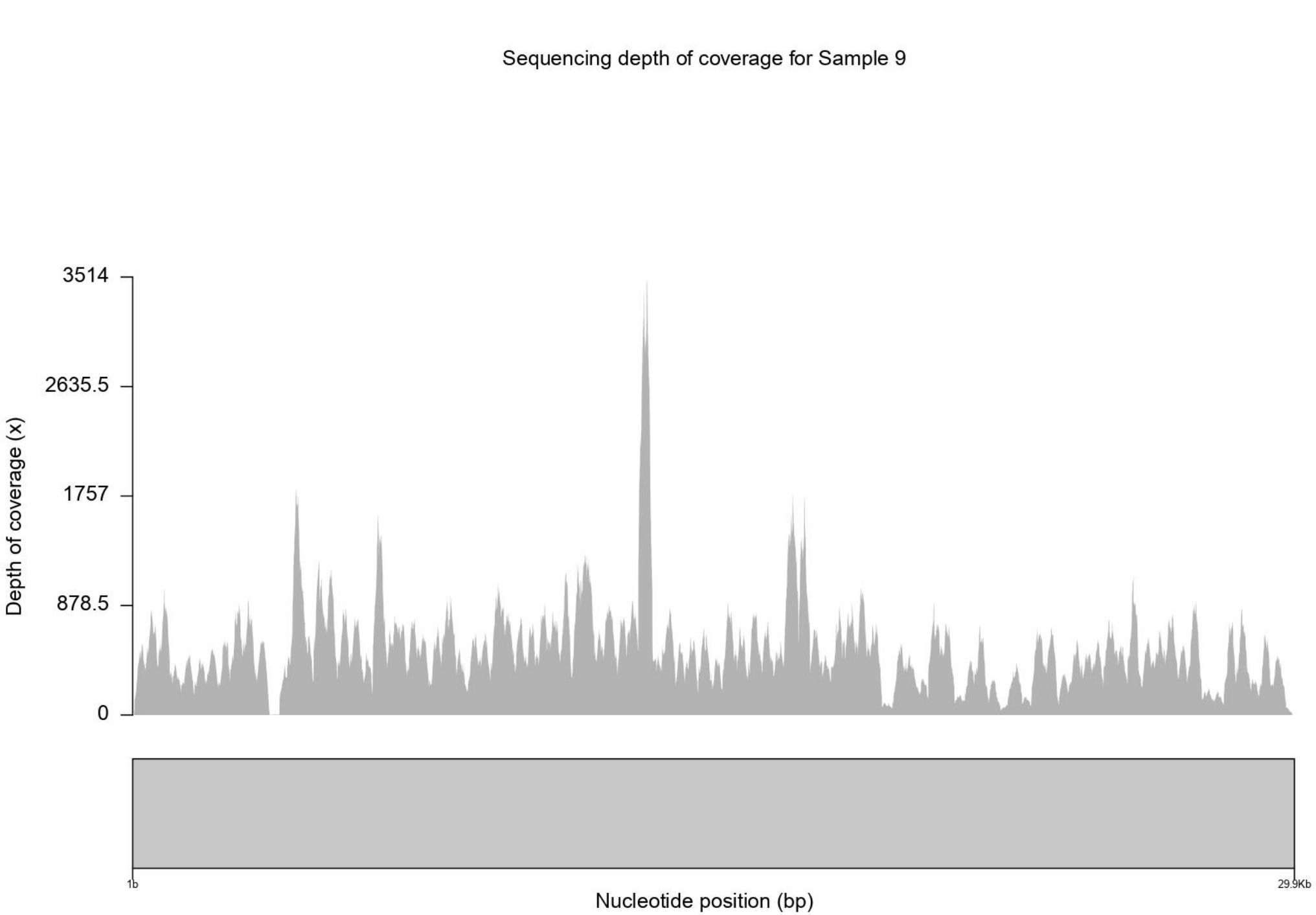

Sequencing depth of coverage for Sample 4
2291
1718.25
1145.5
572.75
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)

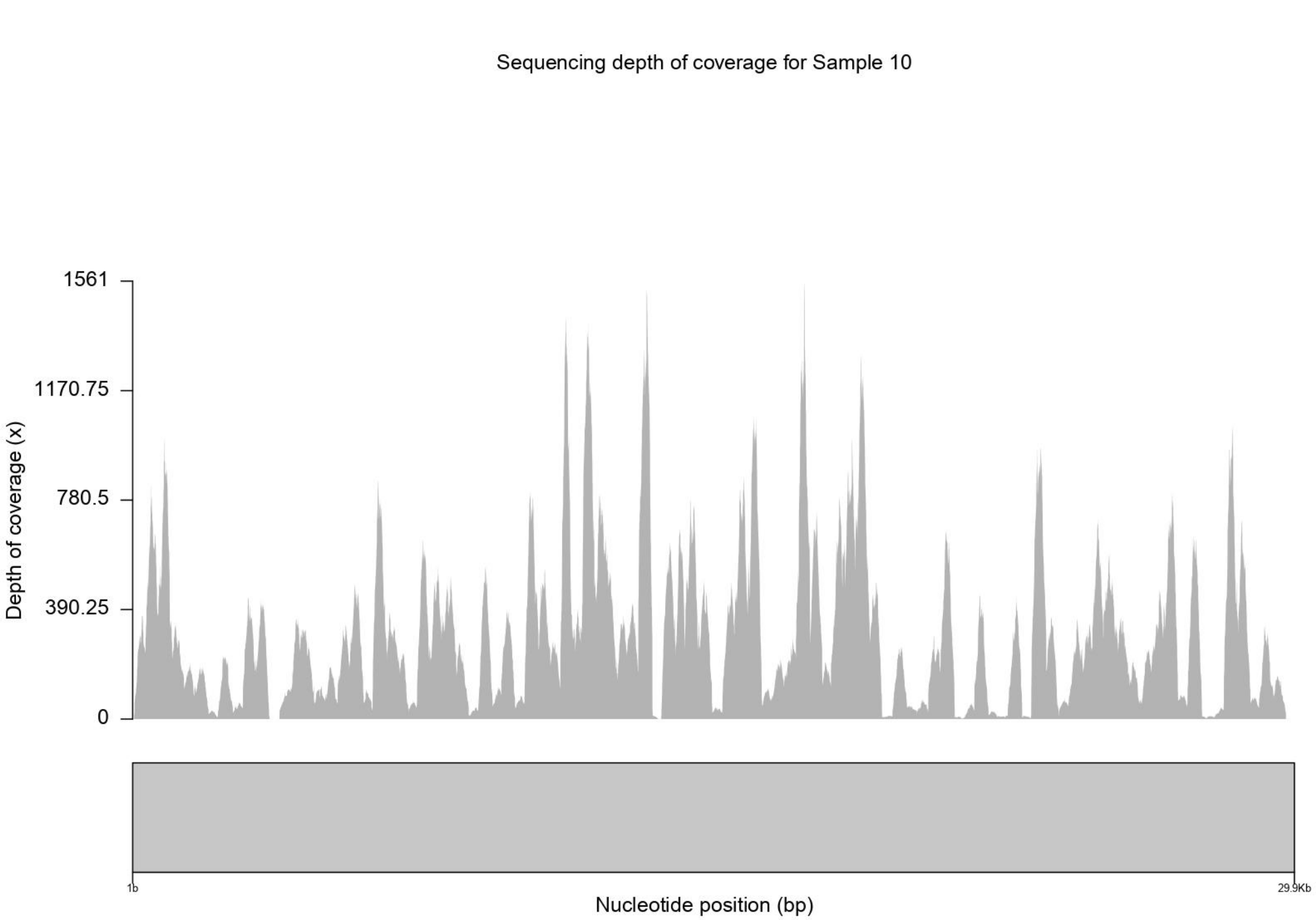

Sequencing depth of coverage for Sample 5
281
210.75
140.5
70.25
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)

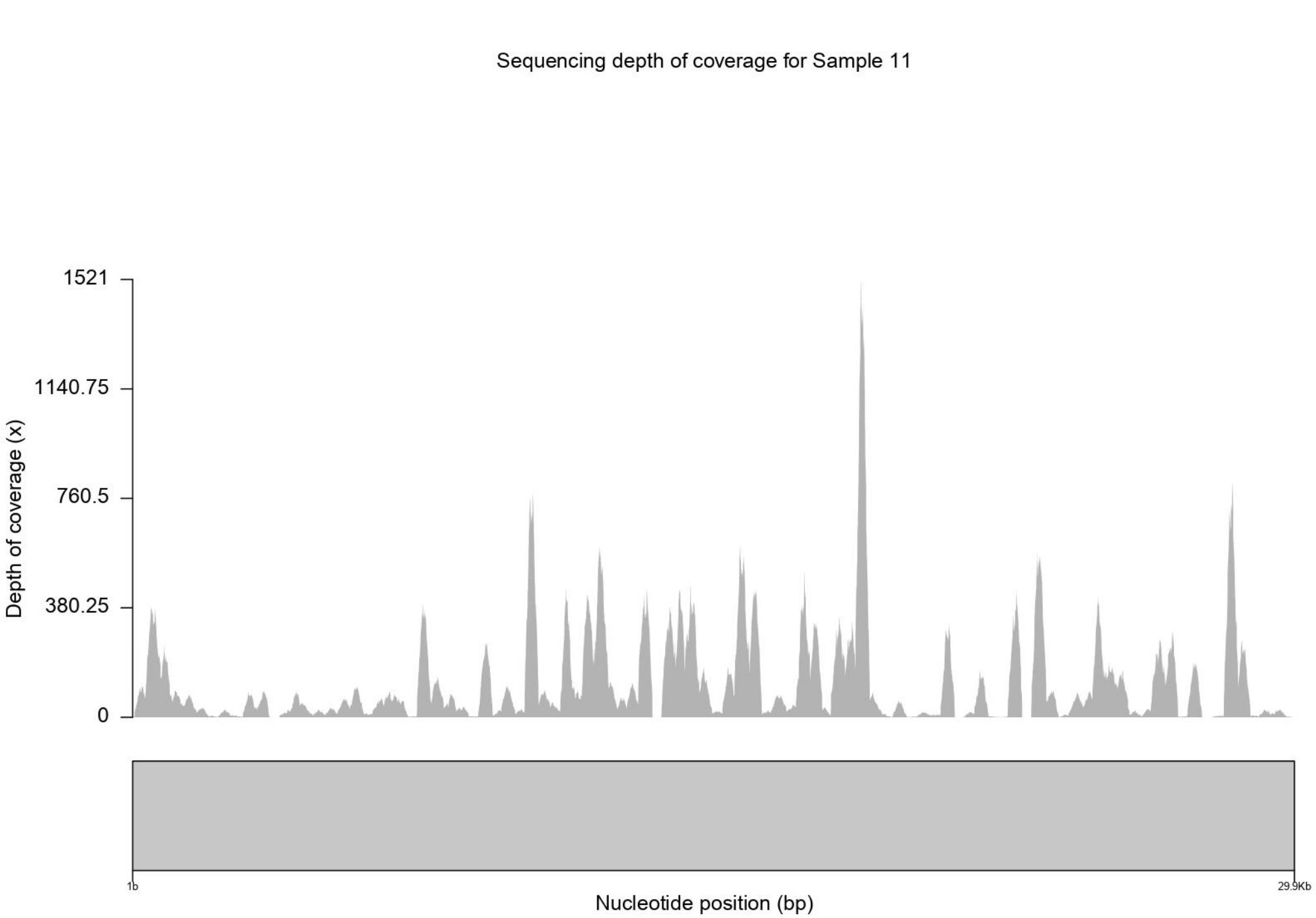

Sequencing depth of coverage for Sample 6
1965
1473.75
982.5
491.25
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)

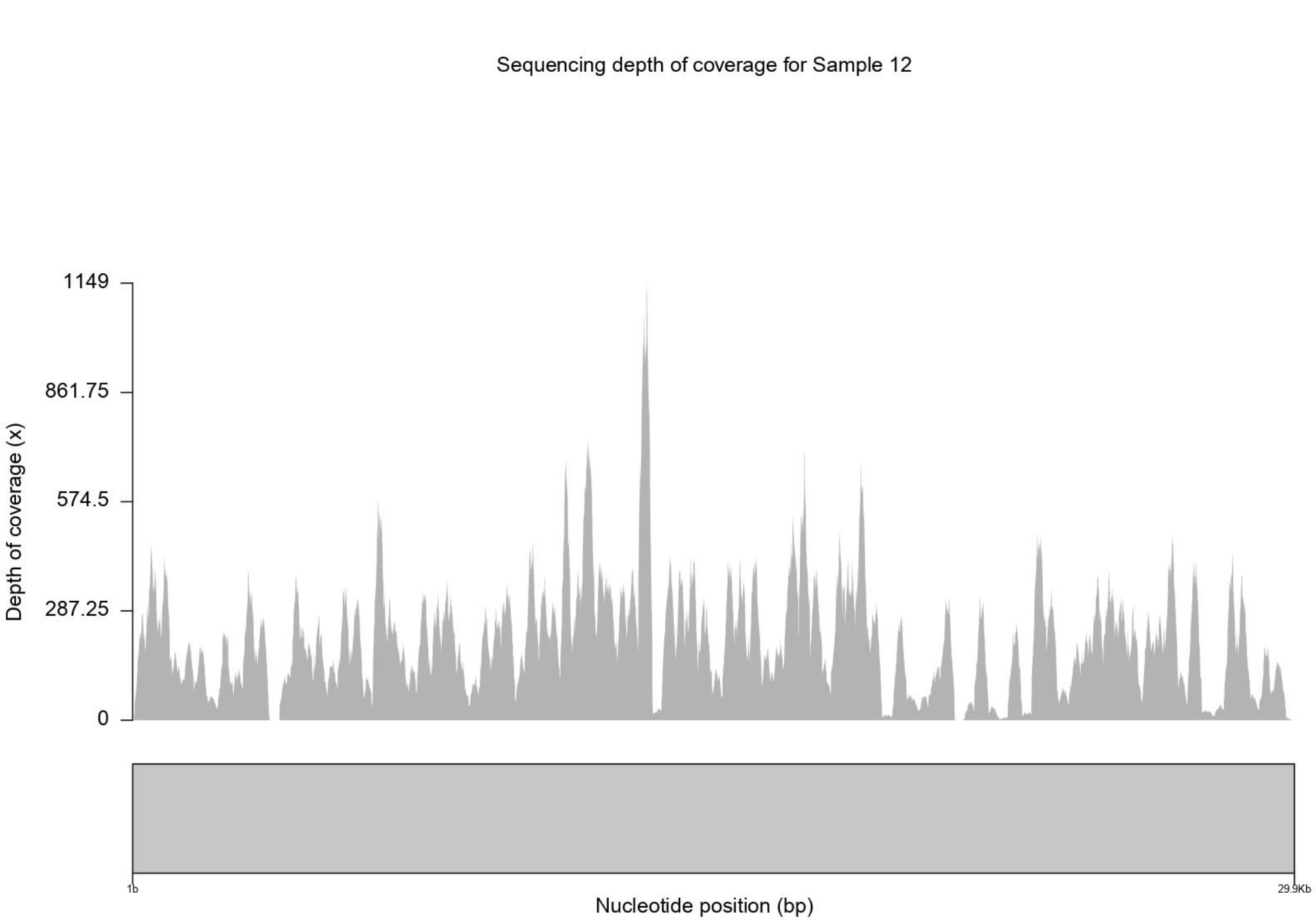

Sequencing depth of coverage for Sample 7
2733
2049.75
1366.5
683.25
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)


Sequencing depth of coverage for Sample 8
1461
1095.75
730.5
365.25
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)


Sequencing depth of coverage for Sample 9
3514
2635.5
1757
878.5
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)


Sequencing depth of coverage for Sample 10
1561
1170.75
780.5
390.25
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)


Sequencing depth of coverage for Sample 11
1521
1140.75
760.5
380.25
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)


Sequencing depth of coverage for Sample 12
1149
861.75
574.5
287.25
0
Depth of coverage (x)
1b
29.9Kb
Nucleotide position (bp)

APÊNDICE II

As tabelas presentes no Apêndice II constituem o segundo material suplementar do artigo que compõe o Capítulo I desta dissertação e compreende todos os sítios das proteínas estruturais E, M e N sob seleção positiva e negativa nos genomas de SARS-CoV-2 do estado do Rio Grande do Sul, de acordo com os testes realizados através dos métodos Hyphy, FUBAR, FEL e SLAC.

Protein E

Codon	FUBAR			FEL				SLAC		
	Alpha	Beta	Post. prob.	Alpha	Beta	LRT	Prob.	dS	dN	Prob
7	-	-	-	-	-	-	-	0.000	4.194	0.047
9	-	-	-	0.000	18.119	3.843	0.0500	-	-	-
23	-	-	-	-	-	-	-	2.368	10.333	0.018
30	-	-	-	-	-	-	-	0.000	3.904	0.047
55	-	-	-	-	-	-	-	4.363	9.713	0.096
61	-	-	-	-	-	-	-	4.468	10.156	0.089
66	-	-	-	-	-	-	-	0.000	3.563	0.098
68	-	-	-	0.000	20.332	3.931	0.0474	-	-	-
72	-	-	-	-	-	-	-	0.000	7.333	0.009
73	-	-	-	0.000	28.091	4.469	0.0345	-	-	-

Post. prob.: Posterior probability. Prob.: Probability

Protein M

Codon	FUBAR			FEL				SLAC		
	Alpha	Beta	Post. prob.	Alpha	Beta	LRT	Prob.	dS	dN	Prob
2	-	-	-	-	-	-	-	0.000	37,014	0.000
3	-	-	-	-	-	-	-	0.000	33.219	0.000
6	-	-	-	0.000	4.650	3.907	0.0481	-	-	-
10	-	-	-	-	-	-	-	22.016	44.162	0.003
15	-	-	-	0.000	6.449	2.810	0.0937	-	-	-
17	-	-	-	0.000	8.960	7.608	0.0058	-	-	-
19	-	-	-	0.000	12.176	3.092	0.0787	-	-	-

28	-	-	-	-	-	-	-	11.136	48.188	0.000
29	-	-	-	-	-	-	-	20.819	45.915	0.001
40	-	-	-	-	-	-	-	19.942	45.908	0.000
53	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2.233**	**0.000**	**0.089**
63	-	-	-	0.000	3.395	3.389	0.0656	22.952	50.757	0.000
68	-	-	-	-	-	-	-	20.355	45.483	0.001
70	-	-	-	-	-	-	-	21.529	45.483	0.001
78	-	-	-	0.000	4.383	3.805	0.0511	-	-	-
82	-	-	-	0.611	9.586	5.085	0.0241	-	-	-
83	-	-	-	-	-	-	-	0.000	35.983	0.000
86	-	-	-	0.000	3.387	2.895	0.0889	0.000	49.412	0.000
87	-	-	-	-	-	-	-	22.507	49.597	0.000
90	-	-	-	-	-	-	-	29.481	49.597	0.000
94	-	-	-	0.000	10.239	8.764	0.0031			
109	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**60.884**	**19.968**	**0.000**
125	-	-	-	0.000	5.193	3.556	0.0593	0.000	35.069	0.000
146	-	-	-	-	-	-	-	21.471	46.930	0.001
148	-	-	-	-	-	-	-	0.000	34.988	0.000
155	-	-	-	-	-	-	-	0.000	35.021	0.000
160	-	-	-	-	-	-	-	0.000	34.909	0.000
162	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**69.781**	**0.000**	**0.000**
193	-	-	-	-	-	-	-	12.885	45.916	0.000
197	-	-	-	-	-	-	-	0.000	33.746	0.000
209	-	-	-	0.000	6.562	4.807	0.0283	-	-	-
211	-	-	-	-	-	-	-	19.906	45.764	0.000

Post. prob.: Posterior probability. Prob.: Probability

Protein N

Codon	FUBAR			FEL				SLAC		
	Alpha	Beta	Post. prob.	Alpha	Beta	LRT	Prob.	dS	dN	Prob
3	-	-	-	0.000	9.490	8.917	0.0028	-	-	-
6	-	-	-	-	-	-	-	49.756	108.066	0.000
7	-	-	-	-	-	-	-	0.000	19.124	0.000
8	-	-	-	-	-	-	-	0.000	12.994	0.001
9	1.162	6.689	0.9171	0.000	11.856	8.821	0.0030	-	-	-
10	-	-	-	0.000	3.102	3.360	0.0879	-	-	-
11	-	-	-	0.000	3.102	3.360	0.668	-	-	-
12	-	-	-	-	-	-	-	0.000	63.292	0.000
13	-	-	-	-	-	-	-	36.325	81.345	0.000
18	-	-	-	0.000	2.303	3.692	0.0547	-	-	-
19	-	-	-	-	-	-	-	0.000	78.367	0.000
24	-	-	-	0.000	3.438	4.398	0.0360			
26	-	-	-	-	-	-	-	0.000	16.729	0.000
29	-	-	-	-	-	-	-	0.000	27.159	0.000
34	0.627	6.934	0.9840	0.000	2.294	3.605	0.0576	-	-	-
41	-	-	-	-	-	-	-	49.113	25.654	0.001
47	-	-	-	-	-	-	-	0.000	12.578	0.001
55	-	-	-	-	-	-	-	35.684	77.889	0.000
63	0.626	6.643	0.9772	0.000	8.489	11.415	0.0007	0.000	30.509	0.000
80	-	-	-	-	-	-	-	0.000	4.993	0.019
89	-	-	-	-	-	-	-	0.000	9.913	0.001
90	-	-	-	-	-	-	-	36.993	80.225	0.000
103	-	-	-	0.000	2.296	2.899	0.0886	0.000	17.188	0.000

110	**-**	**-**	**-**	**2.022**	**0.000**	**2.844**	**0.0917**	**2.391**	**0.000**	**0.078**
113	-	-	-	-	-	-	-	0.000	6.763	0.001
119	-	-	-	0.000	2.420	3.775	0.0520	-	-	-
122	-	-	-	-	-	-	-	16.149	38.502	0.001
124	-	-	-	-	-	-	-	6.922	26.476	0.000
128	-	-	-	-	-	-	-	0.000	20.433	0.000
131	-	-	-	-	-	-	-	0.000	27.113	0.000
134	-	-	-	-	-	-	-	0.000	21.889	0.000
135	-	-	-	0.000	3.460	4.405	0.0358	-	-	-
137	-	-	-	-	-	-	-	0.000	27.113	0.000
144	-	-	-	-	-	-	-	1.673	43.558	0.000
145	-	-	-	-	-	-	-	0.000	17.678	0.000
151	0.595	7.120	0.9879	0.000	5.180	7.784	0.0053	37.283	81.437	0.000
152	-	-	-	-	-	-	-	43.391	92.002	0.000
155	-	-	-	-	-	-	-	16.880	39.458	0.001
161	-	-	-	-	-	-	-	11.474	27.233	0.004
166	-	-	-	-	-	-	-	0.000	30.658	0.000
172	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**3.599**	**0.000**	**0.025**
177	-	-	-	-	-	-	-	0.000	10.875	0.001
178	-	-	-	0.000	2.524	3.785	0.0517	-	-	-
179	-	-	-	-	-	-	-	11.269	42.388	0.000
182	0.610	4.020	0.9256	-	-	-	-	46.782	101.306	0.000
183	-	-	-	0.000	1.885	3.350	0.0672	-	-	-
185	-	-	-	0.000	2.593	3.879	0.0489	11.101	27.546	0.003
191	-	-	-	-	-	-	-	33.484	76.229	0.000
193	-	-	-	0.000	2.479	2.983	0.0842	-	-	-
199	-	-	-	-	-	-	-	0.000	7.981	0.000
200	-	-	-	0.000	2.556	3.801	0.0512	-	-	-
201	-	-	-	-	-	-	-	0.000	16.667	0.000

202	-	-	-	0.000	4.482	6.323	0.0119	0.000	17.927	0.000
204	-	-	-	0.000	4.375	2.854	0.0911	31.486	64.818	0.000
207	-	-	-	0.000	2.826	4.030	0.0447	-	-	-
208	0.604	4.019	0.9266	0.000	2.751	3.977	0.0461	33.854	74.294	0.000
209	-	-	-	-	-	-	-	1.295	30.662	0.000
210	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**146.488**	**12.188**	**0.000**
211	-	-	-	-	-	-	-	18.024	39.410	0.001
213	-	-	-	-	-	-	-	0.000	7.940	0.015
214	-	-	-	-	-	-	-	7.984	30.524	0.000
215	0.620	3.320	0.9035	0.000	4.566	7.028	0.0080	1.095	5.805	0.058
227	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2.168**	**0.000**	**0.098**
229	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**88.182**	**17.547**	**0.000**
232	-	-	-	-	-	-	-	0.000	57.507	0.000
234	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**139.068**	**34.974**	**0.000**
235	-	-	-	0.000	1.898	3.356	0.0670	-	-	-
236	-	-	-	-	-	-	-	23.614	80.735	0.000
238	0.613	5.716	0.9707	0.000	5.145	7.604	0.0058	28.146	100.807	0.000
239	-	-	-	-	-	-	-	0.000	36.903	0.000
243	-	-	-	-	-	-	-	25.577	87.236	0.000
251	-	-	-	-	-	-	-	4.595	11.189	0.056
252	-	-	-	-	-	-	-	12.006	28.435	0.003
253	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**55.898**	**9.584**	**0.000**
260	-	-	-	-	-	-	-	0.000	21.430	0.000
264	-	-	-	-	-	-	-	11.916	28.534	0.003
265	0.733	5.366	0.9119	-	-	-	-	-	-	-
270	-	-	-	-	-	-	-	0.000	60.055	0.000
271	-	-	-	-	-	-	-	0.000	9.913	0.000
274	**-**	**-**	**-**	**3.080**	**0.000**	**3.973**	**0.0462**	**2.392**	**0.000**	**0.078**
276	-	-	-	-	-	-	-	0.000	10.377	0.001

286	-	-	-	-	-	-	-	8.438	23.000	0.013
289	1.190	6.874	0.9166	0.000	6.378	4.554	0.0328	**176.613**	**40.051**	**0.000**
297	-	-	-	-	-	-	-	0.000	54.035	0.000
300	-	-	-	0.000	5.212	6.131	0.0133	-	-	-
305	-	-	-	-	-	-	-	0.000	4.999	0.018
308	-	-	-	-	-	-	-	16.799	39.588	0.001
317	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**158.639**	**10.087**	**0.000**
324	-	-	-	-	-	-	-	35.069	76.589	0.000
326	-	-	-	0.000	2.827	4.030	0.0447	-	-	-
327	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**3.000**	**0.000**	**0.041**
330	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**301.703**	**53.459**	**0.000**
333	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2.382**	**0.000**	**0.087**
341	-	-	-	-	-	-	-	0.000	30.662	0.000
343	-	-	-	-	-	-	-	0.000	54.035	0.000
360	-	-	-	-	-	-	-	0.000	21.715	0.000
362	1.235	7.839	0.9224	0.000	7.674	4.785	0.0287	0.000	10.862	0.000
365	-	-	-	-	-	-	-	0.000	62.294	0.000
366	1.234	7.818	0.9223	0.000	7.672	4.767	0.0290	0.000	11.841	0.000
368	-	-	-	0.435	8.488	7.846	0.0051	11.010	26.463	0.004
374	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**127.269**	**0.000**	**0.000**
377	-	-	-	0.717	12.533	10.100	0.0015	-	-	-
378	-	-	-	-	-	-	-	0.000	16.816	0.000
379	-	-	-	0.000	3.839	4.554	0.0328	-	-	-
380	-	-	-	-	-	-	-	0.000	13.608	0.001
381	-	-	-	-	-	-	-	34.170	77.480	0.000
383	-	-	-	2.259	8.637	2.848	0.0915	**113.407**	**91.536**	**0.054**
386	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**88.256**	**17.551**	**0.000**
391	0.602	5.675	0.9571	0.000	7.591	9.092	0.0026	-	-	-
399	-	-	-	-	-	-	-	1.687	19.771	0.000

401	-	-	-	0.000	2.543	3.029	0.0818	0.000	28.324	0.000
402	-	-	-	0.713	7.655	5.090	0.0241	-	-	-
412	-	-	-	-	-	-	-	0.000	20.389	0.000
413	-	-	-	0.000	2.462	2.990	0.0838	0.000	27.394	0.000
418	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**155.995**	**32.824**	**0.000**

Post. prob.: Posterior probability. Prob.: Probability

APÊNDICE III

O artigo que constitui o Apêndice III, intitulado “Population-based prevalence surveys during the Covid-19 pandemic: A systematic review” teve como objetivo a revisão de literatura e avaliação da confiabilidade de estudos populacionais que estimassem a prevalência populacional da COVID-19. Com isto, foi possível identificar a qualidade das metodologias e possíveis vieses relacionados a estes estudos.

O trabalho encontra-se publicado na revista Reviews in Medical Virology (https://onlinelibrary.wiley.com/journal/10991654), que possui Fator de Impacto (JCR 2021) = 11.043 e Qualis/CAPES = A1 e seu manuscrito, juntamente de seus materiais suplementares se encontram disponíveis na íntegra *(Open Access)* e podem ser acessados por meio do seguinte endereço: https://doi.org/10.1002/rmv.2200.

APÊNDICE IV

O artigo que constitui o Apêndice IV, intitulado “E484K as an innovate phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil” teve como objetivo analisar o aumento da disseminação da mutação E484K da proteína *spike* em genomas de SARS-CoV-2 brasileiros e avaliar os sítios que estão sob seleção positiva nesta proteína.

O trabalho se encontra publicado na revista Infection, Genetics and Evolution (https://www.sciencedirect.com/journal/infection-genetics-and-evolution), que possui Fator de Impacto (JCR 2021) = 4.393 e Qualis/CAPES = A2 e seu manuscrito, juntamente de seus materiais suplementares, encontram-se disponíveis mediante assinatura e podem ser acessados por meio do seguinte endereço: https://doi.org/10.1016/j.meegid.2021.104941.

APÊNDICE V

O artigo que constitui o Apêndice V, intitulado “Genomic epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil” teve como objetivo estudar a evolução e epidemiologia do SARS-CoV-2 na cidade de Esteio/RS, por meio de análises filogenéticas e filodinâmicas.

O trabalho se encontra publicado na revista BMC Genomics (https://bmcgenomics.biomedcentral.com/), que possui Fator de Impacto (JCR 2021) = 4.547 e Qualis/CAPES = A1 e seu manuscrito, juntamente com seus materiais suplementares, encontram-se disponíveis na íntegra (*Open access*) e podem ser acessados por meio do seguinte endereço: https://doi.org/10.1186/s12864-021-07708-w.