



INSTITUTO LATINO AMERICANO DE
CIÊNCIAS DA VIDA E DA NATUREZA
(ILACVN)
BIOTECNOLOGIA

**ANÁLISE GENÔMICA DAS VARIANTES DE SARS-COV-2
DETECTADAS EM FOZ DO IGUAÇU - PARANÁ/BR E O IMPACTO
DAS MUTAÇÕES *MISSENSE* NAS VIAS DEPENDENTES DE HLA-B**

SAMUEL CHAGAS DE ASSIS

Foz do Iguaçu

2024



INSTITUTO LATINO AMERICANO DE
CIÊNCIAS DA VIDA E DA NATUREZA
(ILACVN)
BIOTECNOLOGIA

**ANÁLISE GENÔMICA DAS VARIANTES DE SARS-COV-2
DETECTADAS EM FOZ DO IGUAÇU - PARANÁ/BR E O IMPACTO
DAS MUTAÇÕES *MISSENSE* NAS VIAS DEPENDENTES DE HLA-B**

SAMUEL CHAGAS DE ASSIS

Trabalho de Conclusão de Curso apresentado ao Instituto Latino Americano de Ciências da Vida e da Natureza (ILACVN) da Universidade Federal da Integração Latin-Americana, como requisito parcial à obtenção do título de Bacharel em Biotecnologia. .

Orientadora: Profa. Maria Cláudia Gross, Ph.D.

Foz do Iguaçu
2024

SAMUEL CHAGAS DE ASSIS

**ANÁLISE GENÔMICA DAS VARIANTES DE SARS-COV-2
DETECTADAS EM FOZ DO IGUAÇU - PARANÁ/BR E O IMPACTO
DAS MUTAÇÕES *MISSENSE* NAS VIAS DEPENDENTES DE HLA-B**

Trabalho de Conclusão de Curso apresentado ao Instituto Latino Americano de Ciências da Vida e da Natureza (ILACVN) da Universidade Federal da Integração Latin-Americana, como requisito parcial à obtenção do título de Bacharel em Biotecnologia. .

BANCA EXAMINADORA

Profa. Maria Cláudia Gross, Ph.D. (Orientadora)
UNILA

Prof. Kelvinson Fernandes Viana, Ph.D.
UNILA

Profa. Maria Leandra Terencio, Ph.D.
UNILA

Foz do Iguaçu
2024

DEDICATÓRIA

Dedico este trabalho à minha família, especialmente, ao Tio Cid (*in memorian*), uma das muitas vidas negligenciadas durante a pandemia de COVID-19.

AGRADECIMENTOS

Minha trajetória acadêmica é um reflexo dos incontáveis esforços que meu pai, Caetano, e minha mãe, Márcia, realizaram para me propiciar a oportunidade de estudar. A eles agradeço as oportunidades, as motivações e os ensinamentos que foram cultivados anos atrás pela Vó Clara, Vô Onélio, Vó Gleci e o Vô Nelson.

O interesse pela ciência e pesquisa que foi cultivada por meio de discussões e ensinamentos compartilhados com minha irmã, Caroline, a quem agradeço por ser minha maior inspiração pessoal e profissional.

A motivação diária para conclusão deste trabalho é fruto dos momentos inspiradores e as conversas de conforto com minha namorada Alexia, a quem agradeço pelo amor e o companheirismo de sempre.

Transformar esses anos de graduação com momentos memoráveis e fazer da minha casa, em Foz do Iguaçu, um verdadeiro lar foi devido àqueles com quem compartilhei cada instante. Ao Felipe (Felps), Anderson (Dylon), Michelli, Maria, Rosane, Matheus, Giulio e Andressa, agradeço por serem tão especiais para mim e pelos cafézinhos, a parceria, o carinho e a paciência.

Por me proporcionar a oportunidade de acessar um conhecimento para além das fronteiras e serem grandes amigos, agradeço ao Eric, Julio, Lucas e Bernie pelo conhecimento e os momentos compartilhados.

A possibilidade de contribuir cientificamente no Laboratório de Pesquisa em Ciências Médicas (LPCM) e abrir portas para outras oportunidades foi graças a orientação da Profa. Dra. Maria Claudia Gross, a quem agradeço pelas conversas e orientação.

A Profa. Dra. Maria Leandra Terencio e Prof. Dr. Carlos Henrique Schneider agradeço o suporte e atenção nas etapas de experimentais do meu projeto de iniciação científica, o qual resultou nos dados experimentais utilizados nesse trabalho.

Aos participantes de cada projeto que acreditaram nas ideias mais (in)consequentes possíveis durante a graduação, agradeço pela parceria e dedicação, especialmente àqueles que confiaram no potencial do Centro Acadêmico de Biotecnologia - UNILA e do Grupo de Biologia Sintética SynFronteras.

Ao Corpo Docente do Curso de Biotecnologia, agradeço a dedicação diária em enfrentar os desafios, de um curso novo em uma universidade nova, no suporte para transformação e formação de novos profissionais.

Agradeço a UNILA pela educação pública, gratuita e de qualidade, além de ressignificar para mim a integração latino-americana no desenvolvimento socioeconômico, cultural e político do povo.

*Es América Latina, la región de las venas abiertas. [...]
La causa nacional latinoamericana es, ante todo, una causa social.*

Eduardo Galeano

Las venas abiertas de América Latina, 1978, p. 14 e 281.

RESUMO

A vigilância genômica do SARS-CoV-2 tem sido fundamental para monitorar o surgimento de mutações e compreender suas implicações na severidade e disseminação do vírus, auxiliando na previsão de impactos na eficácia da vacinação e na resposta imune populacional. Neste estudo computacional, mapeamos as variantes de SARS-CoV-2 circulantes em Foz do Iguaçu, Paraná, Brasil, de 2020 a 2022. Focamos na identificação de mutações *missense*, que alteram a sequência de aminoácidos das proteínas virais, e avaliamos seu efeito na apresentação de epítopes, com ênfase nos alelos HLA-B da população local. Os resultados indicaram que epítopes associados aos alelos B*07:02 e B*27:05 apresentaram maior perda de ligação devido a mutações, e que a dinâmica de algumas variantes pode demonstrar a exportação de variantes ao Paraguai. Embora não tenha sido observada uma correlação direta entre o perfil alélico de pacientes internados em UTI e a circulação de mutações específicas, o estudo revelou a dinâmica mutacional do SARS-CoV-2 ao longo do tempo em Foz do Iguaçu durante a pandemia. Dessa forma, a pesquisa identificou mutações que potencialmente impactaram a disseminação do vírus e ofereceu um panorama das variantes circulantes, contribuindo para futuras estratégias de vigilância epidemiológica na região fronteiriça.

Palavras-chave: COVID-19; variações alélicas; complexo de histocompatibilidade; sequenciamento.

ABSTRACT

Genomic surveillance of SARS-CoV-2 has been essential for monitoring emerging mutations and understanding their implications for viral severity and spread, aiding in predicting impacts on vaccine efficacy and the population's immune response. In this computational study, we mapped SARS-CoV-2 variants circulating in Foz do Iguaçu, Paraná, Brazil, from 2020 to 2022. We focused on identifying missense mutations, which change the viral protein amino acid sequence in an unique position, and assessed their effect on epitope presentationfor HLA-B alleles. The results indicated that epitopes associated with the B*07:02 and B*27:05 alleles showed greater binding loss due to mutations, and that the emergence of certain variants may be related to the city's proximity to Paraguay. Although no direct correlation was found between the allelic profile of ICU-admitted patients and the circulation of specific mutations, the study revealed SARS-CoV-2's mutational dynamics over time in Foz do Iguaçu during the pandemic. Thus, this research identified mutations potentially impacting viral spread and provided an overview of circulating variants, contributing to future epidemiological surveillance strategies.

Keywords: COVID-19; allelic variance; major histocompatibility complex; sequencing.

LISTA DE ILUSTRAÇÕES

Figura 1 - Processamento dos antígenos para apresentação ao sistema imune. (A) Principal via de geração dos peptídeos para apresentação pelas moléculas HLA Classe I; (B) e Classe II.	15
Figura 2 - Etapas do processo de infecção de SARS-CoV-2, incluindo entrada viral, replicação e transcrição, montagem e liberação.	17
Figura 3 - Resíduos âncoras Leucina e Valina de um peptídeo ALGIGILTV na fenda de HLA-A*02:01.	18
Figura 4 - Representação das metodologias computacionais desenvolvidas para predição de epítópos de células T baseada em (A) <i>motifs</i> , (B) métodos quantitativos heurísticos e (C) redes neurais artificiais.	20
Figura 5 - Etapas do <i>pipeline</i> para análise das mutações <i>missense</i> . Coletas de dados genômicos (A), alinhamento das sequências das mesmas variantes (B), alinhamento das sequências consenso de cada variante com a sequência de referência (C), identificação de mutações <i>missense</i> (D), geração dos peptídeos de referência e mutados (E), predição da afinidade de ligação (F) e análise estatística (G). Quadros vermelhos apresentam dados quantitativos de cada etapa.	24
Figura 6 - Esquema demonstrando o conceito de perda (<i>Loss</i>) e ganho (<i>Gain</i>) de afinidade de ligação dos epítópos de referência e mutados.	26
Figura 7 - Diversidade de linhagens de SARS-CoV-2 em Foz do Iguaçu/PR de Mar/2020 a Dez/2022. (A) Prevalência relativa das linhagens ao longo dos meses, quadro superior apresentando o número de genomas sequenciados, linhas superiores demonstrando o período das ondas de acordo com estudos prévios (BASTOS et al., 2021; GIOVANETTI et al., 2022; GRÄF et al., 2024; MOURA et al., 2022); (B) perfil alélico de pacientes do HMPGL obtido a partir do estudo experimental, quadro superior apresentando o número de amostras sequenciadas. Mês correspondente a data de detecção/admissão.	29
Figura 8 - Número de mutações únicas que levaram ao efeito de (<i>Weak</i>) <i>Gain</i> ou (<i>Weak</i>) <i>Loss</i> para (A) cada <i>ORF</i> , (B) cada variante e (C) cada alelo. A contagem considera os efeitos para cada alelo predito. Figura B possui as variantes por ordem da primeira detecção na cidade de Foz de Iguaçu de baixo para cima.	31

LISTA DE TABELAS

Tabela 1 - Comparações entre as Frequências dos Grupos Alélicos.	30
Tabela 2 - Mutações <i>Loss</i> epítopos associados ao alelo B*07:02 e B*27:05	36

LISTA DE ABREVIATURAS E SIGLAS

ACE2	Angiotensin-converting enzyme 2 (Enzima Conversora de Angiotensina 2)
FC	Fold change (Razão da expressão relativa)
FCS	Furin cleavage site (Sítio de Clivagem da Furina)
GISAID	Global Initiative on Sharing All Influenza Data
HLA-B	Human Leukocyte Antigen Class I - B (Antígeno Leucocitário Humano Classe I - B)
HMPGL	Hospital Municipal Padre Germano Lauck
MHC	Major Histocompatibility Complex (Complexo Principal de Histocompatibilidade)
RBD	Receptor Binding Domain (Domínio de Ligação ao Receptor)
ORF	Open Reading Frame
VOC	Variant of Concern (Variante de Preocupação)
VOI	Variant of Interest (Variante de Interesse)

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO BIBLIOGRÁFICA	14
2.1	O Complexo Principal de Histocompatibilidade e o Sistema de Antígeno Leucocitário Humano (HLA)	14
2.2	Estrutura e Apresentação à Células T	15
2.3	Estrutura e Processo de Infecção de SARS-CoV-2	16
2.4	NetMHCpan: Predição de afinidade de ligação entre MHC e antígenos virais	18
3	OBJETIVOS	21
3.1	Geral	21
3.2	Específicos	21
4	MATERIAIS E MÉTODOS	22
4.1	Estudo Experimental	22
4.2	Aspectos Éticos	22
4.3	Seleção de indivíduos COVID-19 positivo	23
4.4	Fase laboratorial	23
4.5	Estudo computacional	23
4.6	Coleta de dados genômicos	24
4.7	Identificação de mutações <i>missense</i> de SARS-CoV-2	25
4.8	Predição da afinidade de ligação de peptídeos virais mutados	25
4.9	Análise estatística	26
4.10	Modelo Estrutural	27
4.11	Ferramentas Computacionais	27
5	RESULTADOS E DISCUSSÃO	28
5.1	Diversidade genômica de SARS-CoV-2 em Foz do Iguaçu - Paraná/BR	28
5.2	Diversidade de mutações <i>missense</i> concentra-se em <i>ORF1ab</i> e <i>spike</i>	30
5.3	Epítopes de B*07:02 apresentam maior perda na afinidade de ligação com mutações <i>missense</i>	35
5.4	Implicações e limitações do estudo na região de fronteira	38
6	CONCLUSÃO	41
	REFERÊNCIAS	42

1 INTRODUÇÃO

A vigilância genômica das variantes circulantes de SARS-CoV-2 permitiu o acompanhamento do surgimento de mutações que pudessem gerar impactos na severidade e circulação do vírus. No entanto, mutações neutras ou moderadamente deletérias foram observadas em grande parte da amostragem presente no banco de dados *Global Initiative on Sharing All Influenza Data* (GISaid). Esse fator se deve à tendência de mutações com grande efeito fenotípico serem minoria comparado às toleradas modificações de aminoácidos de baixo- ou sem-efeito (HARVEY et al., 2021).

Porém, essa pequena minoria de modificações pode contribuir para o aumento do *fitness* e conferir determinadas vantagens em diferentes ciclos virais, como: patogenicidade, infectividade, transmissibilidade e/ou antigenicidade. Nesse sentido, ao longo da pandemia surgiram evidências que alterações fenotípicas das variantes virais estavam diretamente relacionadas com alterações no mecanismo de resposta (neutralização por anticorpos) e reconhecimento (ligação pelos complexos de apresentação de antígeno) do sistema imune (HARVEY et al., 2021; PONTAROTTI; PAGANINI, 2022).

Portanto, esforços foram concentrados em compreender essas diferenças de resposta imune para avaliar o comportamento evolutivo viral, como também, compreender as regiões e marcadores que podem afetar a eficácia de uma vacinação em massa. Em termos de reconhecimento do sistema imune, diferentes trabalhos mapearam epítópos de SARS-CoV-2, a partir da linhagem de origem em Wuhan, China, que são potencialmente reconhecidos por células T (GRIFONI et al., 2020; KIYOTANI et al., 2020).

Essa diversificação do quadro de apresentação de antígenos baseado nas informações mutacionais foi explorada por Hamelin et al. (2022), onde se avaliou o impacto das mutações prevalentes, detectadas na vigilância genômica global de SARS-CoV-2, demonstrando que a identificação de determinadas mutações e grupos alélicos que apresentam influência na resposta imune, baseada na apresentação de epítópos, se torna um indicador de que a monitoramento genômico viral e a identificação do perfil HLA populacional podem convergir para avaliar o quadro de incidência da doença em determinada região e no desenvolvimento de novas vacinas.

Em relação a regiões de fronteira, esses indicadores se tornam mais difíceis de medir e avaliar. O trânsito comercial e urbano podem gerar perturbações em relação a vigilância epidemiológica de cada país, devido a circulação de novas variantes e a diversificação genômica presente na população de fronteira. Portanto, o presente trabalho busca mapear as variantes que potencialmente circularam no município de Foz do Iguaçu - Paraná/BR e a influência das mutações *missense* - uma mutação pontual que codifica um aminoácido não-sinônimo - na apresentação de epítópos em relação ao perfil alélico de HLA-B da região.

2 REVISÃO BIBLIOGRÁFICA

2.1 O Complexo Principal de Histocompatibilidade e o Sistema de Antígeno Leucocitário Humano (HLA)

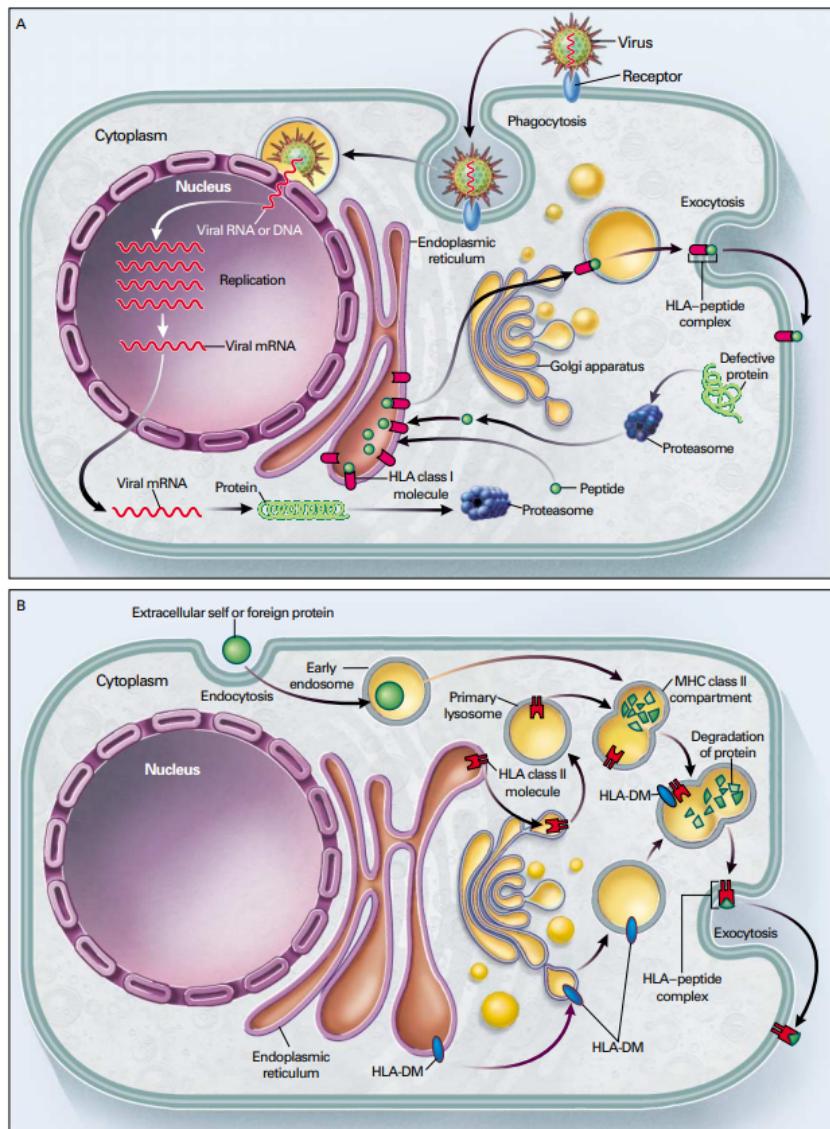
O Complexo Maior de Histocompatibilidade (*Major Histocompatibility Complex* - MHC, em inglês) exerce um papel fundamental no sistema imune, responsáveis pela apresentação de peptídeos na superfície celular para reconhecimento das células T (WIECZOREK et al., 2017). A versão humana do MHC, chamado de sistema HLA (*Human Leukocyte Antigen*, em inglês) é codificado a partir do cromossomo 6 e estima-se que apresenta cerca de 40 diferentes genes de HLA, extremamente polimórficos, distribuídos em três classes. No entanto, apenas duas classes, I e II, estão envolvidas com a resposta imune e possuem diferentes funções e estruturas (KLEIN; SATO, 2000).

Os genes da Classe I são consideravelmente expressos em células somáticas e dependem do tipo de tecido, já a Classe II está presente nas células do sistema imune incluindo células B, células T ativadas, macrófagos, células dendríticas, etc (KLEIN; SATO, 2000; SIEGRIST, 2018). As Classes I e II também se dividem na localização celular e para entender seu funcionamento é necessário compreender a dinâmica celular no processamento de peptídeos intra- e extracelulares. No citosol, proteínas defeituosas, indesejadas ou desgastadas são processadas pelo proteossoma a partir do reconhecimento da ubiquitina. Esse processamento gera peptídeos que são transportados, com auxílio da classe de Transportadores Associados com o Processamento de Antígenos (TAP, sigla em inglês), para o retículo endoplasmático e dependendo de sua afinidade de ligação, são reconhecidos pelo MHC Classe I e exportados por meio do Complexo de Golgi até a superfície extracelular (Figura 1A) (KLEIN; SATO, 2000).

Por outro lado, as proteínas adquiridas do espaçamento extracelular são processadas de maneira diferente. A partir da endocitose (ou fagocitose), as proteínas extracelulares são processadas por meio do mecanismo lisossomal com proteínas proteolíticas para formação de endossomos e reciclagem de aminoácidos. Durante o processamento, o MHC Classe II reconhece os peptídeos pré-processados baseados na afinidade de ligação e são exportados para superfície extracelular (Figura 1B) (KLEIN; SATO, 2000).

Apesar de haver uma diferença funcional e de localização entre as duas classes, seu funcionamento não é restrito às suas atribuições previamente descritas. Em alguns casos, peptídeos extracelulares são processados pela Classe I e peptídeos intracelulares são processados pela Classe II (KLEIN; SATO, 2000).

Figura 1 - Processamento dos antígenos para apresentação ao sistema imune. (A) Principal via de geração dos peptídeos para apresentação pelas moléculas HLA Classe I; (B) e Classe II.



Fonte: Klein e Sato (2000)

2.2 Estrutura e Apresentação à Células T

As estruturas da Classe I e II são compostas por dois domínios, com uma única cadeia-a pesada (Classe I) ou com duas cadeias, cadeia- α e cadeia- β (Classe II). Os dois domínios formam uma base folha- β e duas α -hélices no topo, onde são acomodados os peptídeos. Cada cadeia apresenta um domínio de imunoglobulina, enquanto o MHC Classe I é associado a uma cadeia leve de β -2 microglobulina (β 2m). Por fim, hélices transmembrana ancoram os complexos na membrana (WIECZOREK et al., 2017).

Os genes que compreendem as Classes I e II apresentam alto grau de polimorfismo (os mais conhecidos são: Classe I: HLA-A, -B, -C; Classe II: HLA-DR, -DP, -DQ), os quais possuem uma variação alélica que afeta a ligação de peptídeos nos sulcos de ligação e modulam a seleção de peptídeos apresentados na superfície celular pelo MHC. No sulco de ligação, a estrutura helicoidal se liga aos peptídeos de acordo com a (i) formação de ligações de hidrogênio entre as estruturas do MHC e o peptídeo e a (ii) ocupação de pequenos espaços que ancoram resíduos da cadeia lateral peptídica (WIECZOREK et al., 2017). Apenas alguns desses pequenos espaços já podem determinar a especificidade de cada peptídeo em se ligar ao complexo e geralmente é definido pela conformação dos diferentes alelos (KLEIN; SATO, 2000).

Essa interação das cadeias laterais de cada peptídeo depende da geometria, distribuição de carga e hidrofobicidade na ligação ao sulco. No entanto, a predição de afinidade nessa interação MHC-antígeno não depende somente das características moleculares, como o tamanho (8-12 resíduos para Classe I e 13-25 resíduos para Classe II) ou composição das cadeias laterais, se não das conformações transitórias e energéticas das estruturas do MHC durante suas flutuações de equilíbrio da estrutura molecular. Por isso, o carregamento do MHC é influenciado por diferentes fatores, como disponibilidade de antígenos, atividade das proteases, disponibilidade de chaperonas e interação com proteínas auxiliares, como a Tapasina (Classe I) e HLA-DM (Classe II) (WIECZOREK et al., 2017).

A partir da ligação, o MHC é transportado até a superfície celular para apresentação do antígeno para células T. Na região dos linfonodos, as células T CD4+ reconhecem os抗ígenos apresentados pela Classe II, enquanto as células T CD8+ reconhecem o complexo MHC-peptídeos formados pela Classe I. O reconhecimento dos peptídeos, geralmente, não são o suficiente para ativação das células T, as quais permanecem anérgicas ou se tornam tolerantes na falta de um estímulo secundário (SIEGRIST, 2018).

2.3 Estrutura e Processo de Infecção de SARS-CoV-2

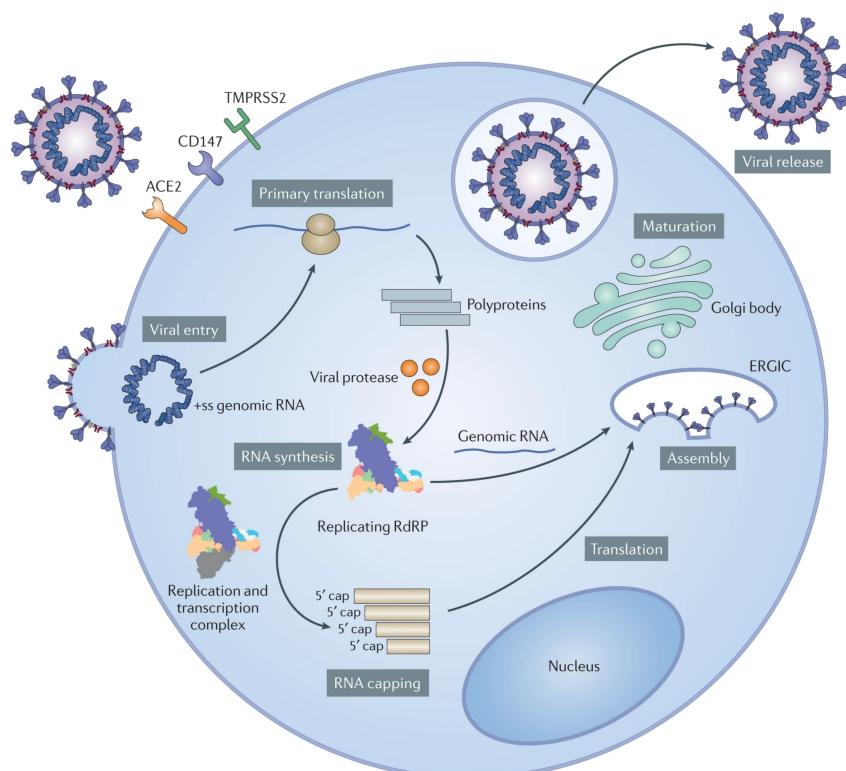
O genoma de SARS-CoV-2 é composto por 14 ORFs, os quais codificam cerca de 29 proteínas virais. Com aproximadamente 30kb, esse genoma contém inicialmente duas ORFs (ORF1a e ORF1b) responsáveis pela expressão de duas poliproteínas (pp1a e pp1ab) que podem ser processadas em 15-16 proteínas não-estruturais; quatro conjunto de proteínas estruturais (*nucleocapsídeo* (N), *spike* (S), *membrana* (M) e *envelope* (E)); e uma série de proteínas acessórias (ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8b, ORF9b e ORF14) (TAY et al., 2020; V'KOVSKI et al., 2021; YANG; RAO, 2021).

Na região terminal 5' estão presentes os genes das poliproteínas, essas proteínas são codificados por duas regiões *ORF1* sobrepostas e são digeridas por duas proteases virais, PL2 e 3CL-pro, resultando em 16 proteínas não-estruturais (NSPs, em inglês), as quais

compõe o complexo de replicação e transcrição viral (RTC). Enquanto a região terminal 3' codifica proteínas estruturas e ORFs intercaladas codificam proteínas acessórias com diferentes funções no processo de infecção e montagem do vírion resultante (VKOVSKI et al., 2021).

O vírus acessa o espaço intracelular por meio de interações entre a proteína S com receptores celulares e proteínas auxiliares (como a Enzima Conversora da Angiotensina 2 - ACE2 e a proteína transmembrana TMPRSS2, respectivamente), acarretando na fusão entre a membrana celular do hospedeiro e a viral (Figura 2). A fusão permite a liberação do RNA viral genômico no espaço intracelular que são traduzidas por ribossomos funcionais sequestrados do mecanismo natural da célula hospedeira. Inicialmente, as poliproteínas pp1a e pp1ab são co-expresas e digeridas em NPSs para se agregarem e formarem o RTC (TAY et al., 2020; YANG; RAO, 2021).

Figura 2 - Etapas do processo de infecção de SARS-CoV-2, incluindo entrada viral, replicação e transcrição, montagem e liberação.



Fonte: Yang e Rao (2021)

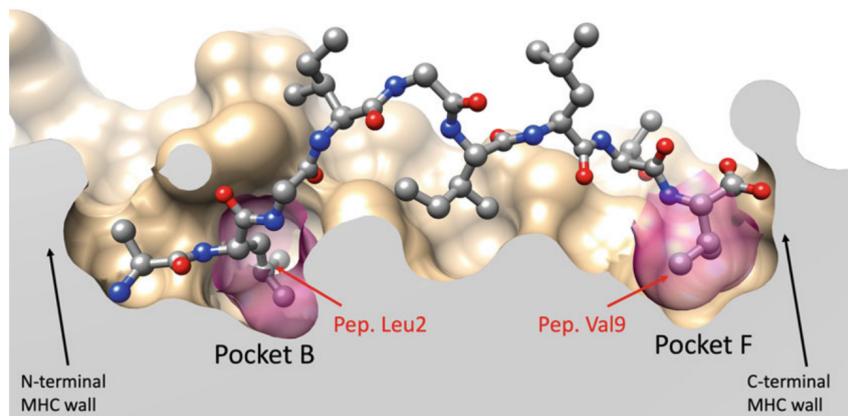
Durante o ciclo da infecção viral, algumas organelas se auto-replicam a partir do retículo endoplasmático (RE) e propiciam um ambiente favorável para replicação e transcrição de mRNAs subgenômicos (*sg-mRNAs*). Aparentemente, esse mecanismo de

replicação é evitado no citosol, provavelmente, para não ser detectado por sensores do sistema imune inato (V'KOVSKI et al., 2021). As proteínas estruturais que são traduzidas no RE acabam sendo translocadas entre as membranas do RE até o Complexo de Golgi por meio de um compartimento intermediário, onde o vírion resultante é formado. Por fim, o vírion formado é secretado por exocitose para seguir com o ciclo de infecção nas demais células do hospedeiro (V'KOVSKI et al., 2021; YANG; RAO, 2021).

2.4 NetMHCpan: Predição de afinidade de ligação entre MHC e antígenos virais

A identificação de epítópos de células T é um grande desafio devido à significativa variação em seu reconhecimento entre indivíduos. Os genes envolvidos com o sistema HLA são os mais diversos do genoma humano, além dos fatores ambientais que afetam o histórico de interação entre o sistema de apresentação de抗ígenos. As moléculas de MHC apresentam especificidades de ligação distintas, caracterizando um sistema de apresentação de epítópos robusto e variado. Na Figura 3 é possível verificar a quantidade de interações que podem ser geradas a partir da ligação entre o MHC e o peptídeo.

Figura 3 - Resíduos âncoras Leucina e Valina de um peptídeo ALGIGILTV na fenda de HLA-A*02:01.



Fonte: Perez et al. (2022)

A ligação de peptídeos ao MHC é uma etapa necessária para que o sistema imune adaptativo e inato possa ser ativado, esses ligantes são classificados como epítópos de células T. Essa etapa representa o processo mais seletivo da via de reconhecimento de抗ígenos e compreender os mecanismos que influenciam esse evento contribui para o desenvolvimento de ferramentas, as quais podem predizer potenciais alvos que ativam ou

não uma resposta imune por meio de vias dependentes de MHC (PETERS; NIELSEN; SETTE, 2020).

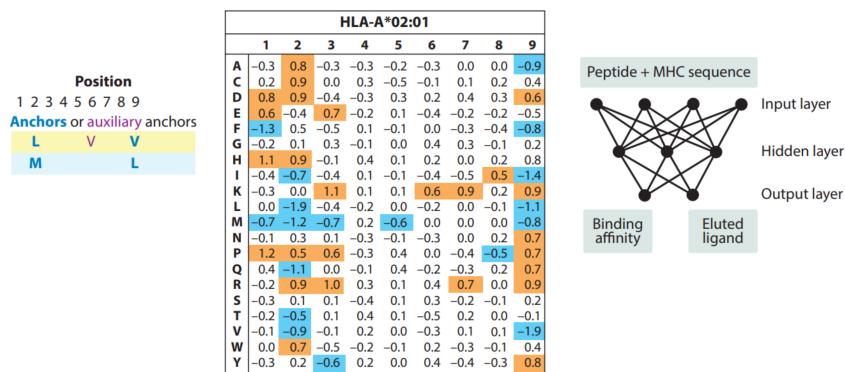
Essas aplicações corroboram com os avanços no desenvolvimento de vacinas, imunoterapia do câncer e pesquisa de doenças infecciosas. Por isso, múltiplos esforços são aplicados no desenvolvimento de métodos computacionais capazes de acuradamente prever a ligação de peptídeos e mapear os epítópos ao MHC-I e MHC-II. Entre eles, existem o SYFPEITHI (RAMMENSEE et al., 1999), netMHCpan (REYNISSON et al., 2020), MHCflurry (O'DONNELL; RUBINSTEYN; LASERSON, 2020), MHCnuggets (SHAO et al., 2020) e MixMHCpred (GFELLER et al., 2023).

De acordo com Peters, Nielsen e Sette (2020), esses métodos computacionais são responsáveis por mapear os epítópos de células T que evoluíram ao longo do tempo, adicionando características a cada versão. Inicialmente, visava-se identificar os *motifs* formados pelos aminoácidos dos peptídeos, os quais correspondiam aos resíduos de ligação ao MHC. Determinadas posições desses peptídeos permitem um limitado número de substituintes e com propriedades de cadeia-laterais semelhantes, os quais mantêm o potencial de ligação MHC-ligante. Esses resíduos foram chamados de resíduos âncoras e outros influentes foram classificados como auxiliares, os quais também possuíam um espaçamento e especificidade determinante para afinidade de ligação que quando somados caracterizavam os motifs dos resíduos ligantes ao MHC (Figura 4A).

Em um segundo momento, esses métodos foram aprimorados para abordagens quantitativas da afinidade de ligação do MHC utilizando abordagens de aprendizado de máquina. Nesse sentido, uma abordagem inicial foi determinar pontuações, a partir de valores heurísticos, para os vinte aminoácidos em cada posição que quantitativamente representasse o seu impacto esperado. Esses valores eram estimados a partir de dados experimentais e eram derivados para se obter uma pontuação final de uma sequência peptídica em relação à determinada variante de MHC. A partir da eficiência do método heurístico, abordagens baseadas em aprendizado de máquina supervisionado foram implementadas (Figura 4B).

Em uma terceira abordagem, as atuais ferramentas utilizam arquiteturas de redes neurais artificiais especializadas para integrar informações das interações lineares e não-lineares a partir da sequência do MHC e dados da ligação MHC-ligante. No entanto, dados experimentais são um limitante para o treinamento supervisionado, dado o número de variantes existentes. Essa limitação motivou o desenvolvimento de técnicas que baseiam-se na construção de matrizes virtuais que caracterizam o perfil de ligação de determinada variante de MHC, sem dados experimentais, por meio da comparação dos resíduos do sulco de ligação do MHC com os resíduos do sulco de ligação de variantes que foram avaliados experimentalmente. Essa última abordagem ficou conhecida como *pan-specific*, exibindo alta confiabilidade na predição de epítópos para variantes MHC sem dados experimentais (Figura 4C).

Figura 4 - Representação das metodologias computacionais desenvolvidas para predição de epítopes de células T baseada em (A) *motifs*, (B) métodos quantitativos heurísticos e (C) redes neurais artificiais.



Fonte: Peters, Nielsen e Sette (2020)

O primeiro método computacional a implementar predições *pan-specific* em redes neurais para moléculas do MHC Classe I foi o NetMHCpan (HOOF et al., 2009). Esse método complementa a informação sobre a ligação MHC-ligante e ligante eluído no treinamento do modelo preditivo com informações sobre os aminoácidos que definem o sulco de ligação do MHC. Além disso, a partir de aprimoramentos, o NetMHCpan-4.1 permite desambiguar peptídeos que tiveram baixa resolução de anotação para determinados alelos a partir de dados experimentais e aumento a performance de predição em relação à demais ferramentas (REYNISSON et al., 2020).

Portanto, nesse trabalho, busca-se implementar a ferramenta NetMHCpan-4.1 na avaliação dos efeitos na ligação entre peptídeo-MHC a partir dos dados de mutações *missense* de variantes de SARS-CoV-2 identificadas em Foz do Iguaçu/PR.

3 OBJETIVOS

3.1 Geral

Avaliar a relação da diversidade de variantes SARS-CoV-2 que circularam em Foz do Iguaçu/PR, entre 2020 e 2022, e o impacto das mutações *missense* na afinidade de ligação dos epítopos aos complexos de apresentação de antígeno HLA-B.

3.2 Específicos

- Identificar os sítios de mutações *missense* mais relevantes a partir do alinhamento das variantes identificadas entre 2020 e 2022 no GISAID;
- Coletar os grupos alélicos de HLA-B mais frequentes no Estado do Paraná via *Allele Frequency Net Database* e estudos prévios;
- Estimar a afinidade de ligação dos peptídeos de referência e mutados em relação aos alelos HLA-B selecionados utilizando NetMHCpan-4.1.
- Relacionar as mutações *missense*, o período de circulação da variante e o impacto na apresentação de epítopos baseado nos dados preditos.

4 MATERIAIS E MÉTODOS

Os dados experimentais foram extraídos de projetos prévios e descritos nesse trabalho. A análise das mutações *missense* das variantes de SARS-CoV-2 utilizou-se de abordagens computacionais e dados públicos. As etapas computacionais de coletas das sequências genômicas, identificação das mutações e predição de epitopos foram baseadas em metodologias prévias (HAMELIN et al., 2022).

4.1 Estudo Experimental

O município de Foz do Iguaçu, no sul do Brasil, faz fronteira com Paraguai e Argentina. O município tem 258.823 habitantes, uma densidade de 414,58 pessoas por km², urbanização superior a 99% e um Índice de Desenvolvimento Humano de 0,751. A rede de saúde inclui 36 unidades, sendo quatro hospitais, organizados em cinco distritos. O primeiro caso de COVID-19 foi registrado em 12 de março de 2020, seguido pelo primeiro caso de transmissão local em 25 de março, transmissão comunitária em 7 de abril e o primeiro óbito em 26 de abril. As medidas de distanciamento social foram iniciadas a partir de 15 de março (VIANA et al., 2021).

Os pacientes COVID-19 positivo admitidos na Unidade de Tratamento Intensivo (UTI), no Hospital Municipal Paulo Germano Lauck, entre 2020 e 2021, foram submetidos à análise genotípica de HLA-B. Os indivíduos foram selecionados priorizando o histórico hospitalar e quadro clínico melhor documentados durante o período de internação. A pesquisa fez parte da Ação 9 de enfrentamento a COVID-19 no âmbito da Universidade Federal da Integração Latino-Americana – Unila, intitulada “Medicina personalizada para tratamento de pacientes COVID-19 em Foz do Iguaçu” (PORTARIA No 193/2020/GR) (ProjCOVID).

4.2 Aspectos Éticos

Essa etapa faz parte do projeto ”Perfil da população do oeste paranaense acometido de Síndrome Respiratória Aguda Grave entre 2020 a 2022”, aprovado pelo Comitê de Ética em Pesquisa Envolvendo Seres Humanos, CAAE 36189220.3.0000.8527, em 2020, Número do Parecer: 4.250.900.

4.3 Seleção de indivíduos COVID-19 positivo

As amostras foram coletadas de pacientes admitidos e apresentando testes RT-qPCR positivos anteriormente ou na admissão hospitalar. A coleta foi realizada a partir de julho/2020 à novembro/2020 e maio/2021 à julho/2021. Os pacientes com declaração de nacionalidade estrangeira, tempo de internação no Hospital superior à 65 dias (considerados *outliers*) e que apresentam resultados ambíguos para os genótipos dos grupos alélicos de HLA foram removidos desta análise.

O histórico hospitalar foi consultado por meio do Sistema de Gestão Tasy (Koninklijke Philips N.V., Inc., Amsterdam, NL) utilizado pelo Hospital Municipal Padre Germano Lauck, assim como, os quadros clínicos atuais foram coletados no projeto Proj-COVID.

4.4 Fase laboratorial

De cada paciente foram coletados cerca de 6 ml de sangue periférico em tubo contendo anticoagulante EDTA (Tubo Greiner Bio-One) e armazenados em refrigeração - 80°C pela equipe de enfermagem hospitalar. As amostras tiveram o DNA extraído via kit extração *Geno Plus Mini VIOGENE*, numerados e armazenados no banco de amostras do LPCM-UNILA . A quantificação de cada amostra foi realizada em espectrofotômetro NanoDrop ND2000 (Thermo Fischer Inc., Waltham, MA, EUA).

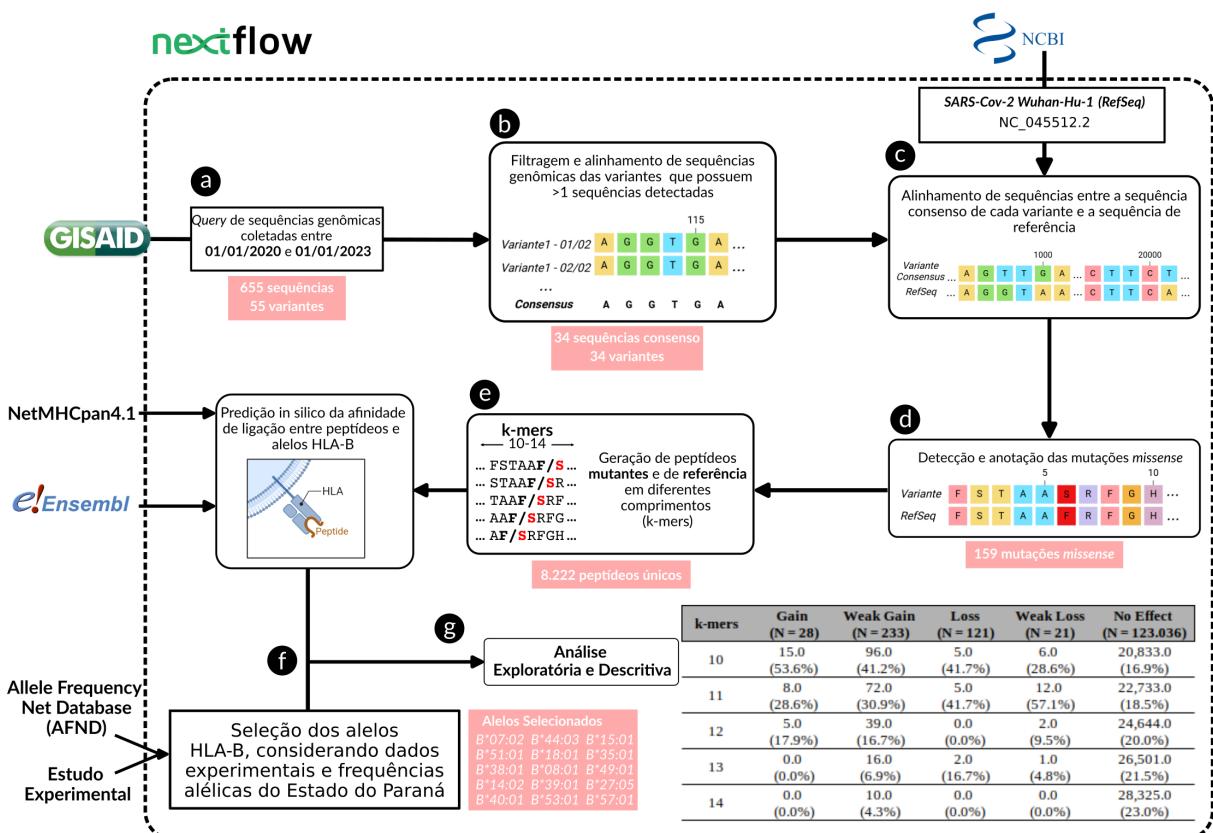
Os exons de HLA classe I *loci* B foram genotipados por Sequenciamento de Grupo Específico (GSSP) via kit *SeScore Sequencing GSSP* (One Lambda, Inc), a partir de amostras do DNA extraído dos pacientes SARS-CoV-2 positivo. As amostras foram checadas, purificadas e sequenciadas em sequenciador automático de *DNA ABI 3500* (Thermo Fischer Inc., Waltham, MA, EUA). As análises dos dados obtidos foram realizadas em Software uTYPE Dx HLA *Sequence Analysis* (One Lambda, Inc). Eventualmente algumas amostras apresentaram resultados com ambiguidade em baixo nível de resolução (grupos alélicos), essas amostras foram filtradas.

4.5 Estudo computacional

Os métodos computacionais buscaram coletar os dados genômicos de variantes que circularam no município de Foz do Iguaçu (Paraná, Brasil) (Figura 5A e 5B), identificação de mutações missense (Figura 5C e 5D), predição da afinidade de ligação de peptídeos virais mutados (Figura 5E e 5F) e análise descritiva e estatística (Figura 5G). Além disso, dados experimentais de projetos anteriores foram utilizados para direcionar a escolha dos

alelos HLA-B selecionados neste estudo.

Figura 5 - Etapas do *pipeline* para análise das mutações *missense*. Coletas de dados genômicos (A), alinhamento das sequências das mesmas variantes (B), alinhamento das sequências consenso de cada variante com a sequência de referência (C), identificação de mutações *missense* (D), geração dos peptídeos de referência e mutados (E), predição da afinidade de ligação (F) e análise estatística (G). Quadros vermelhos apresentam dados quantitativos de cada etapa.



Fonte: O Autor (2024)

4.6 Coleta de dados genômicos

As 655 sequências genômicas de SARS-CoV-2 foram coletadas via GISAIDR (versão 0.9.10) (WIRTH; DUCHENE, 2022) a partir do banco de dados EpiCoV, pertencente ao *Global Initiative on Sharing All Influenza Data* (GISAID) (KHARE et al., 2021). A requisição foi realizada para sequências genômicas completas ($\geq 29\text{kb}$) coletadas entre 01/01/2020 e 01/01/2023, extraídas a partir de amostras provenientes da cidade de Foz do Iguaçu - Paraná/Brasil (Figura 5A e 5B). As 55 variantes detectadas foram inferidas com base na classificação fornecida pelo *EpiCoV*. Sequências com mais de 5% das bases desconhecidas, extraídas de amostras ambientais ou variantes que tiveram apenas uma

sequência coletada foram excluídas. O banco de sequências utilizado neste estudo pode ser acessado via gisaid.org/EPI_SET_240517bd.

4.7 Identificação de mutações *missense* de SARS-CoV-2

As sequências consenso foram obtidas a partir de sequências alinhadas de mesma variante. A sequência de referência SARS-CoV-2 Wuhan-1 (NC_045512.2) foi alinhada com cada sequência consenso individualmente utilizando *minimap2 v2.26-r1175*. As sequências mapeadas foram ordenadas e convertidas em um arquivo *.bam* utilizando *samtools v1.18*. A chamada das variantes foi realizada utilizando *bcftools mpileup v1.17* no modo haplóide e mutações somáticas foram anotadas a partir de *snpEff v4.3* com a função *ann* (Figura 5D).

4.8 Predição da afinidade de ligação de peptídeos virais mutados

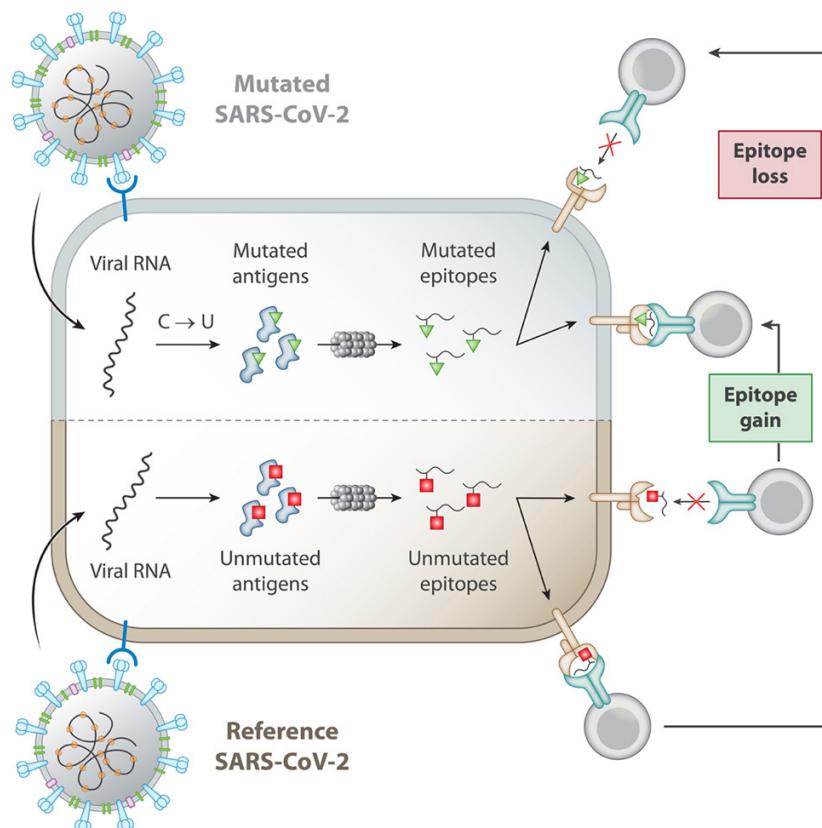
Os peptídeos virais foram selecionados a partir da posição de mutação, gerando *k-mers* de 10 à 14 aminoácidos com a presença de mutação única em cada uma das posições, assim como, peptídeos não mutados como referência (Figura 5E). Os peptídeos que apresentavam similaridade e identidade completa com peptídeos do genoma humano *GRCh38* foram descartados utilizando o *Ensembl BioMart* (HARRISON et al., 2024; KINSELLA et al., 2011). A afinidade de ligação dos peptídeos virais selecionados foi predita contra 15 alelos de HLA-B (B*07:02, B*08:01, B*14:02, B*15:01, B*18:01, B*27:05, B*35:01, B*38:01, B*39:01, B*40:01, B*44:03, B*49:01, B*51:01, B*53:01, B*57:01). A escolha dos alelos teve como base os grupos alélicos presentes nas coletas experimentais realizadas em 2020 e 2021 pelo laboratório, além de acrescentar alelos frequentes no Estado Paraná detectados no Registro Brasileiro de Doadores Voluntários de Medula Óssea (REDOME) e em estudos prévios (Apêndice A).

A predição da afinidade de ligação dos peptídeos foi realizada pelo NetMHCpan 4.1 EL, esse método utiliza uma rede neural treinada com dados experimentais de afinidade de ligação e dados de ligantes eluídos para gerar uma pontuação que indica a probabilidade de um peptídeo ser um ligante para os tipos de HLA especificados (REYNISSON et al., 2020). A probabilidade de ligação é expressa em forma de classificação percentílica (%rank), onde os ligantes classificados como fracos (*Weak binder*, WB) obtiveram um %rank abaixo de 2.0, enquanto os ligantes fortes (*Strong binder*, SB) obtiveram um %rank abaixo de 0.5 ou não ligantes (*Non binder*, NB) obtêm um %rank acima de 2.0 (Figura 5E).

Com base nesse sistema de classificação, os pares de peptídeos referência/mutados que apresentaram uma transição de classificação a partir das mutações foram reclassifi-

cados com base na perda ou ganho de afinidade de ligação. As mutações que fizeram peptídeos passar de ligantes fortes (SB) para ligantes fracos (WB) ou não-ligantes (NB) levam a classificação de uma leve perda (Weak loss) ou uma perda de ligação (Loss), respectivamente. Enquanto que mutações que causam um ganho de afinidade de ligação de WB para SB ou de NB para SB foram consideradas como um leve ganho de ligação *Weak gain* ou *Gain*, respectivamente (Figura 5F e 6).

Figura 6 - Esquema demonstrando o conceito de perda (*Loss*) e ganho (*Gain*) de afinidade de ligação dos epítópos de referência e mutados.



Fonte: Hamelin et al. (2022)

4.9 Análise estatística

O teste não-paramétrico de Sinais de Wilcoxon foi aplicado para definir a significância estatística entre os valores de %rank para peptídeos de referência e mutantes em relação a cada alelo analisado (Figura 5F). A amplitude de impacto de cada mutação que gerou perda na afinidade de ligação foi analisada por meio da equação (1), transformação \log a partir do valor de %rank com mutação normalizado em relação ao %rank de referência:

$$\log_2(Fold\ Change) = \log_2\left(\frac{\%rank_{mutação}}{\%rank_{referência}}\right)$$

Os grupos alélicos detectados no Estudo Experimental foram organizados pela data de admissão hospitalar e grupos alélicos que foram detectados em menos de 10 pacientes, entre 2020 e 2021, foram agrupados em "Outros". Os p-valores das comparações das frequências de HLA-B entre os grupos foram obtidos por Teste Exato de Fisher ao nível de significância adotado de 5% ($p < 0.5$). Quando mencionado, os p-valores foram ajustados para múltiplas comparações com o método de Bonferroni.

4.10 Modelo Estrutural

O modelo estrutural da glicoproteína spike (na forma clivada no sítio de clivagem da furina) com uma ligação ao ACE2 foi utilizado a partir dos conjuntos de coordenadas do PDB ID:7A94 (BENTON et al., 2020). O software Pymol (The PyMOL Molecular Graphics System, v.2.2.0) foi empregado para visualização.

4.11 Ferramentas Computacionais

O *pipeline* foi construído com Nextflow v23.10.1.5891 e as diferentes etapas para predição da afinidade de ligação foram executadas via Python 3.10, utilizando um script adaptado a partir do *pipeline nf-core/epitopeprediction v2.2.1* (MOHR et al., 2024). As análises de dados e gráficos foram produzidos no software RStudio v2023.06.02 (R versão 4.4). Os códigos e os dados estão disponíveis em github.com/chagas98/sarscovHLAFoz.

5 RESULTADOS E DISCUSSÃO

5.1 Diversidade genômica de SARS-CoV-2 em Foz do Iguaçu - Paraná/BR

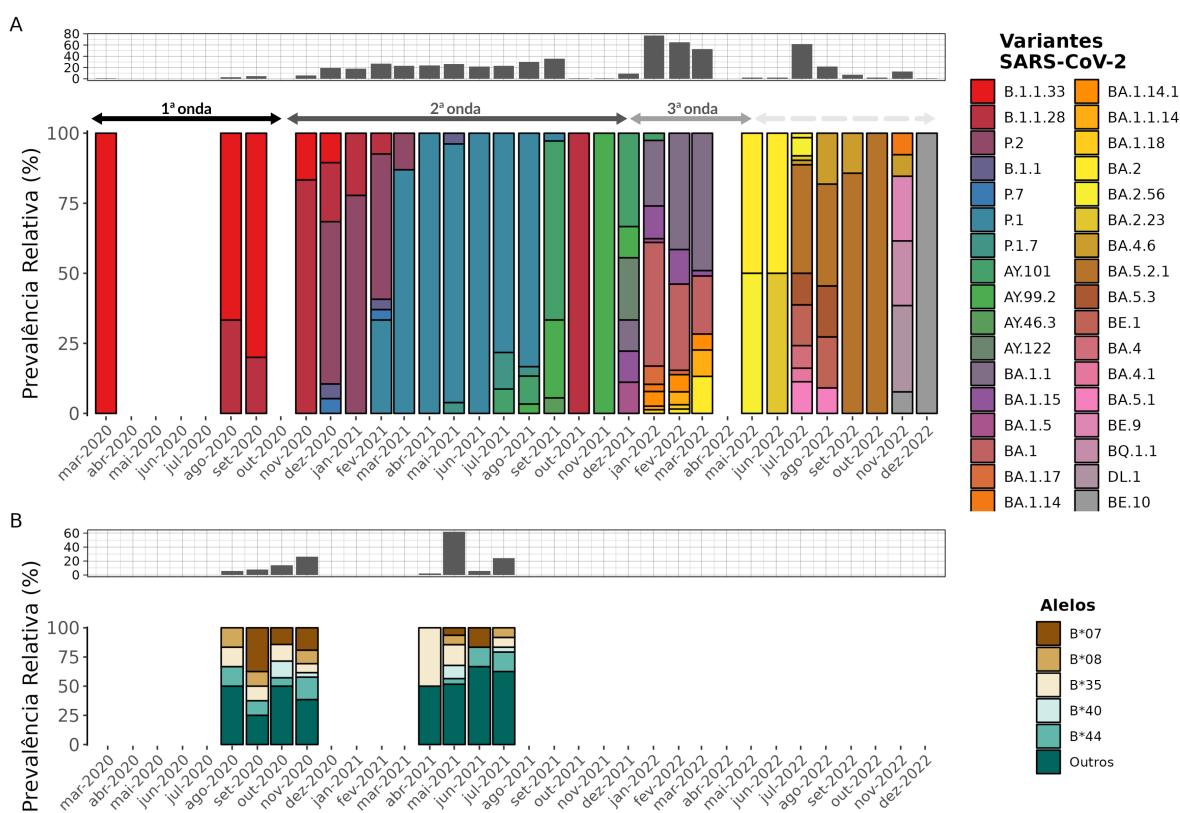
No total, excluindo sequências de amostras ambientais, 640 sequências foram utilizadas nas análises. De acordo com a classificação realizada pelo GISAID, foram detectadas 34 variantes em Foz do Iguaçu/PR no período entre 01/01/2020 e 01/01/2023. A quantidade de genomas sequenciados foram 9 (1.5%), 265 (45.68%), 197 (33.96%) e 109 (18.7%) para a primeira onda (Fev/2020 a Out/2020) , segunda onda (Nov/2020 a Dez/2021), terceira onda (Jan/2022 a Mai/2022) e o segundo semestre de 2022, respectivamente (Figura 7A).

Essa diversidade genômica contribui para compreender a prevalência das mutações durante as ondas epidemiológicas de SARS-CoV-2. Na cidade da tríplice fronteira, a primeira onda foi marcada pela presença das variantes B.1.1.28 e B.1.1.33. A segunda onda foi marcada pela predominância das variantes P.2 (Zeta), P.1 (Gamma), P.1.7 (Gamma) e sub-linhagens Delta (AY.101, AY.99.2, AY.46.3 e AY.122), respectivamente. A terceira onda (Jan/2022 a Mai/2022) possuiu uma maior diversidade de variantes em relação às duas primeiras fases, sendo marcada majoritariamente pela circulação de sub-linhagens Ômicron/BA.1.* e BA.2.*. No segundo semestre de 2022, houve uma predominância da variante Ômicron/BA.5.2.1 em relação às demais variantes, assim como, uma diminuição no número de genomas sequenciados.

As variantes Gamma e Zeta surgiram durante a segunda onda, com a variante Gamma emergindo ao final do mês de Novembro/2020, no município de Manaus/AM. Enquanto a variante Zeta foi descrita em Outubro/2020, no município do Rio de Janeiro/RJ. No entanto, estudos anteriores apresentam indícios do surgimento da variante Zeta no estado do Paraná ao final de Agosto/2020, tendo detecções originadas em diferentes localidades da região Sul e Sudeste (FARIA et al., 2021; GIOVANETTI et al., 2022). Essa constatação é similar com o comportamento epidemiológico observado no município de Foz do Iguaçu, onde a Zeta foi detectada dois meses antes da Gamma (Figura 7A).

Ao final da segunda onda, as variantes Delta foram detectadas majoritariamente. Com exceção da Venezuela e Guianas, países da América do Sul não tiveram uma ressurgência com essas variantes (GRÄF et al., 2024). No entanto, sua prevalência no sul brasileiro foi notória em relação às demais linhagens, tendo indícios do surgimento da AY.101 no Estado do Paraná (ARANTES et al., 2022).

Figura 7 - Diversidade de linhagens de SARS-CoV-2 em Foz do Iguaçu/PR de Mar/2020 a Dez/2022. (A) Prevalência relativa das linhagens ao longo dos meses, quadro superior apresentando o número de genomas sequenciados, linhas superiores demonstrando o período das ondas de acordo com estudos prévios (BASTOS et al., 2021; GIOVANETTI et al., 2022; GRÄF et al., 2024; MOURA et al., 2022); (B) perfil alélico de pacientes do HMPGL obtido a partir do estudo experimental, quadro superior apresentando o número de amostras sequenciadas. Mês correspondente a data de detecção/admissão.



Fonte: O Autor (2024)

A composição das variantes das duas primeiras ondas foi comparada de forma indireta com o perfil alélico de pacientes admitidos no Hospital Municipal de Foz do Iguaçu. O estudo experimental obteve o resultado de 74 amostras de pacientes COVID-19, sendo 27 coletados em 2020 e 47 coletados em 2021. O perfil alélico foi obtido para os genes de HLA-B em uma resolução baixa a nível de grupos alélicos, amostras ou resultados com ambiguidades não foram tratados. Alelos que apresentaram um número total de ocorrências inferior a dez foram agrupados em “Outros” (Figura 7B).

O B*07 apresentou diferença significativa entre os anos de coleta ($p = 0.02$) e óbitos ($p = 0.004$) em 2020 e 2021, esse último sugerindo uma mudança na ocorrência em pacientes com alta e óbito (Tabela 1). Já o B*41 apresentou diferença entre os desfechos

Tabela 1 - Comparações entre as Frequências dos Grupos Alélicos.

Alelos	<i>p</i>	<i>p.adj</i>	G1 ¹	G2 ²	G1(n)	G2(n)	Outros 1 (n)	Outros 2 (n)
B*41	0.052	1.0	Alta20	Óbito20	5	0	23	26
B*07	0.020	0.47	2021	2020	5	10	89	44
B*07	0.004	0.1	Óbito20	Óbito21	7	1	19	41

¹Grupo 1²Grupo 2

Fonte: O Autor (2024)

para 2020, mas não atingiu o limiar de significância estatística. Quando ajustado os p-valores para múltiplas comparações, no entanto, os p-valores perdem significância. O grupo alélico B*35 apresentou diferença entre os desfechos de cada ano, mas não foi detectada significância estatística.

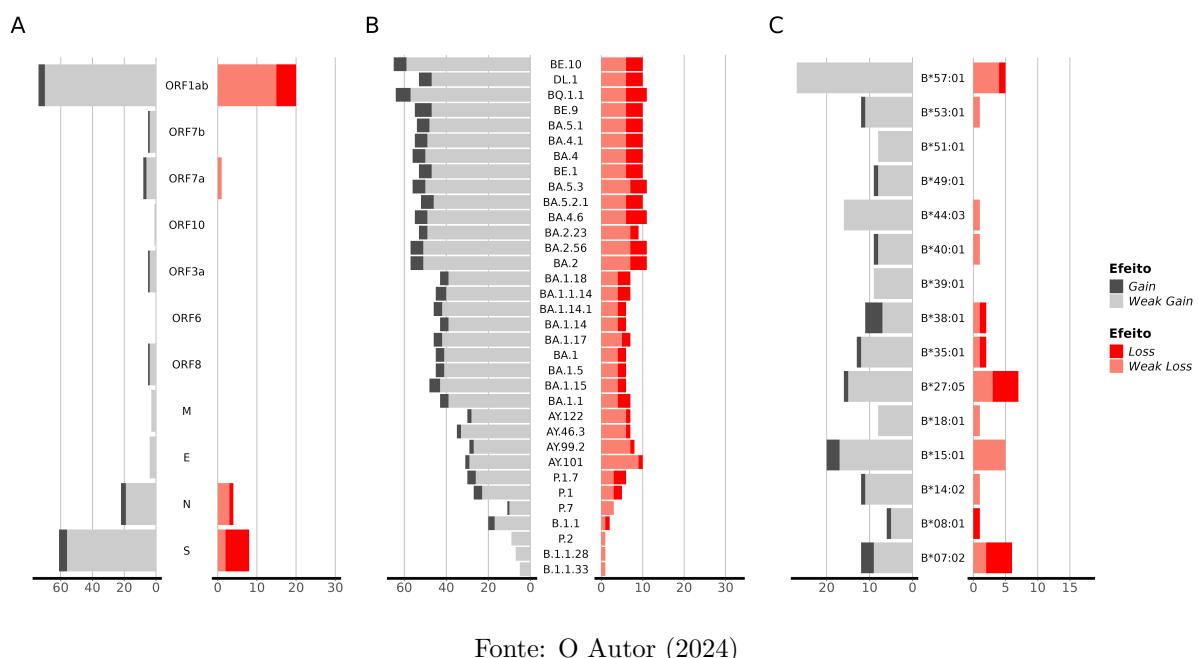
Além disso, quando comparadas as frequências relativas com as frequências cole-tadas do REDOME, o grupo alélico B*07 apresentou uma diferença percentual evidente entre 2020 (18%), 2021 (5%) e REDOME (7%).

5.2 Diversidade de mutações *missense* concentra-se em *ORF1ab* e *spike*

A partir das sequências genômicas, foram gerados 8.222 peptídeos entre *10mers* e *14mers* proveniente das posições de 185 mutações únicas identificadas. As regiões com maior número de mutações *missenses* foi a *ORF1ab* (58), *spike* - *S* (56) e *Nucleocapsídeo* - *N* (20), as regiões restantes tiveram menos de 10 mutações. No entanto, a glicoproteína *spike* e as poliproteínas *ORF1ab* apresentaram maior número de mutações classificadas em *Loss*, 4 e 3, respectivamente. Enquanto a proteína *N* apresentou uma única mutação com forte perda (*Loss*) de afinidade de ligação (Figura 8A).

Além disso, a afinidade de ligação foi mensurada para cada agrupamento. No entanto, como os epíticos não mutados são previstos como não ligantes, essas mutações não foram comparadas com a lista de epíticos validados com alta afinidade de ligação. Ainda é necessário determinar experimentalmente se as mutações do SARS-CoV-2 que aumentam a afinidade dos epíticos ao HLA podem melhorar as respostas das células T e auxiliar no controle do vírus em pacientes com COVID-19.

Figura 8 - Número de mutações únicas que levaram ao efeito de (*Weak*) *Gain* ou (*Weak*) *Loss* para (A) cada *ORF*, (B) cada variante e (C) cada alelo. A contagem considera os efeitos para cada alelo predito. Figura B possui as variantes por ordem da primeira detecção na cidade de Foz de Iguaçu de baixo para cima.



Fonte: O Autor (2024)

Por outro lado, algumas regiões não possuíram mutações com impacto negativo na previsão de afinidade por epítopos. Por exemplo, o envelope apresentou uma única mutação e que não causou qualquer impacto negativo nas análises de afinidade de ligação, a mutação T9I (posição 9, T → I) prevaleceu ao longo de 23 variantes. Estudos anteriores sugerem que essa mutação pode estar relacionada à ineficiência do canal de íons transmembrana do vírus, reduzindo a cascata de sinalização intracelular, resultando em uma diminuição da produção citocinas e diminuição da virulência. Esses fatores podem ter contribuído pelo seu surgimento em Março/2020 e prevalência ao longo das variantes Alfa, Beta, Gamma e Delta em 0,20%, chegando a 100% de ocorrência nas variantes Ômicron (XIA et al., 2022).

Em Foz do Iguaçu, a mutação T9I possivelmente foi introduzida através das variantes Ômicron (BA.*¹) e tiveram prevalência em variantes posteriores, de BA.1.1 até BE.10. Essa tendência de prevalência pode ser observada na chegada de outras variantes além da Ômicron, as quais contribuíram para o aumento de epítopos mutados que tiveram perda de afinidade de ligação predita, como no caso das variantes Gamma (P.1 + P.1.7), Zeta (P.2), Delta (AY.101 + AY.99.2 + AY.46.3) (Figura 8B).

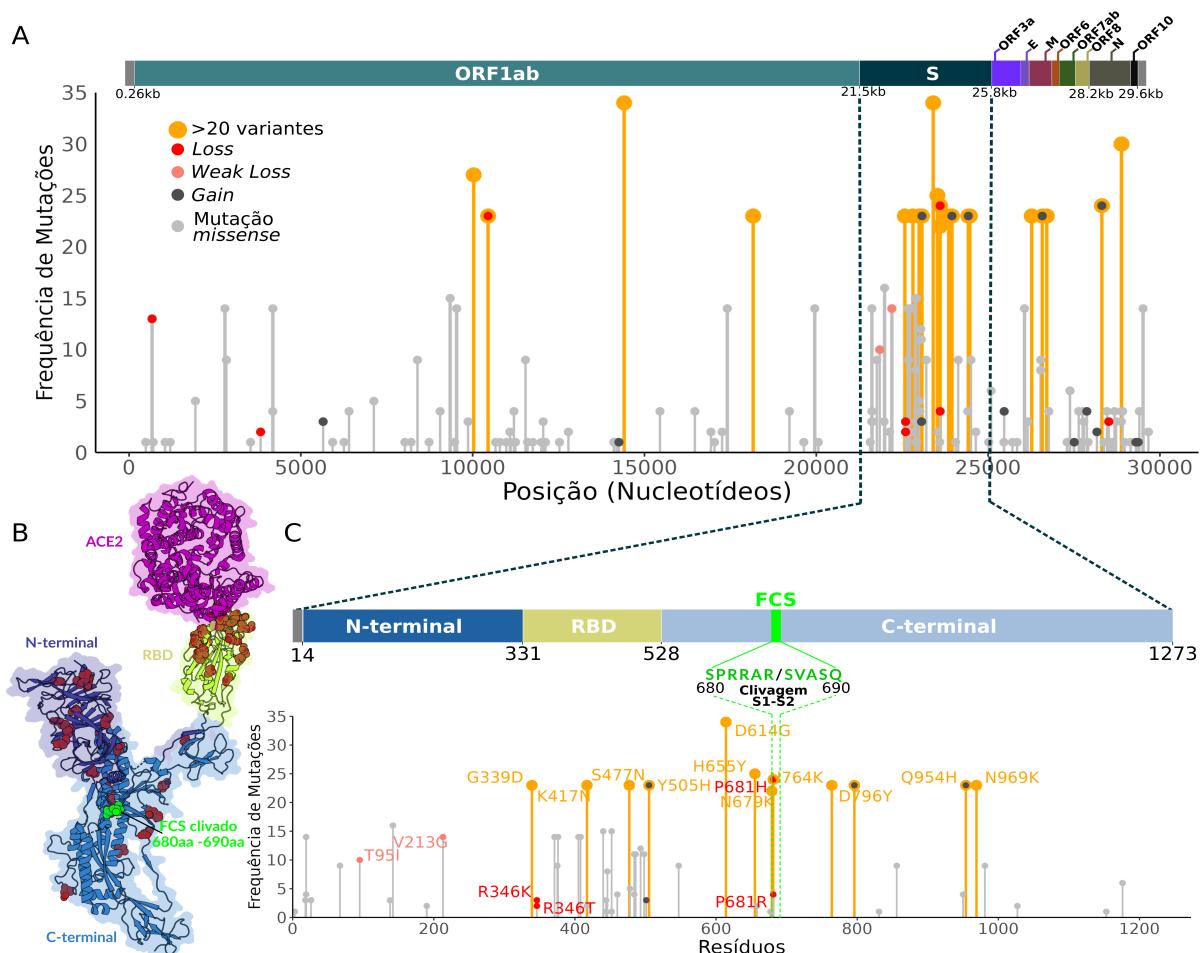
Em *ORF1ab*, as mutações com maior prevalência são P323L/P4715L (34), as quais são provenientes da mutação 14408C>U à nível de nucleotídeo. Esta mutação resulta em uma mutação *missense* P4715L a nível de aminoácidos da poliproteína *ORF1ab* e

aparece como mutação P323L na enzima RNA-polimerase RNA-dependente (RdRp). Essa mutação teve uma perda suave na afinidade de ligação em relação ao alelos B*15:01, em conformidade com os relatos da associação dessa mutação com casos severos e elevação na taxa de mortalidade (TOYOSHIMA et al., 2020).

Outras mutações com frequência significativa em *ORF1ab*, como S135R (13) e P132H/P3395H (23) foram classificados como *Loss*, com as duas últimas mutações levando uma perda de epítocos B*07:02. Esse fato corrobora com análises de bioinformática prévias, as quais sinalizam a perda de afinidade de epítocos em relação aos alelos HLA-B7 *supertype* (HAMELIN et al., 2022). Na Figura 8C, é possível observar uma tendência dos alelos B*07:02 e B*27:02, principais representantes de HLA-B7 e HLA-B27, na perda de afinidade de ligação a partir de epítocos mutados, contendo o maior número de mutações classificadas como *Loss*, respectivamente.

Em relação à glicoproteína *spike*, a mutação D614G apresentou maior frequência em relação a todas as mutações detectadas na cidade de Foz do Iguaçu, ao total foram 34 variantes portadoras desta mutação (Figura 9A). A D614G não afetou a afinidade de ligação a partir das previsões realizadas para os alelos de HLA-B, apesar da sua ocorrência estar ligada ao aumento da taxa de mortalidade, aumento do sinal de detecção de RT-qPCR devido a carga viral e previsões relacionadas aos alelos HLA-A (HAMELIN et al., 2022; KORBER et al., 2020; TOYOSHIMA et al., 2020). Por outro lado, epítocos com mutações P681H (24), P681R (4), R346K (3) e R346T (2) apresentaram uma perda significativa na afinidade de ligação aos alelos B*27:05 e B*07:02 para as posições 681 e 346, respectivamente. Enquanto as mutações T95I (10) e V213G (14) foram classificadas como *Weak Loss* em relação à B*57:01 e B*38:01, respectivamente.

Figura 9 - Distribuição da posição dos epítopos e suas variantes mutadas dos antígenos de SARS-CoV-2 que circularam em Foz do Iguaçu/PR. (A) Posição das mutações ao longo do genoma de SARS-CoV-2, mutações que ocorreram em mais de 20 variantes estão destacadas em amarelo, mutações que foram classificadas como *Gain*, *Loss* e *Weak Loss* estão destacadas em cinza escuro, vermelho e salmão, respectivamente; (B) estrutura de um monômero da *spike* ligado à ACE2. As posições das mutações e região do FCS (680-690) estão destacadas em esferas vermelha e verde, respectivamente.



Fonte: O Autor (2024)

A mutação P681R é exclusivamente carregada por variantes Delta, já a mutação P681H surge a partir de variantes Ômicron em conjunto com mutação N679K (Figura 9C). Essas mutações ocorrem em uma posição adjacente ao Sítio de Clivagem da Furina (*FCS*, sigla em inglês) e aumentam o número de aminoácidos básicos, característica crucial para o reconhecimento pela furina (HARVEY et al., 2021). Além da perda da prolina que afetam a apresentação de epítopos associados à HLA-B7 supertype, mutações na posição 681 também apresentaram resistência à beta interferon (IFN- β) e conferindo uma evasão em relação ao sistema imune adaptativo e inato (HAMELIN et al., 2022; LISTA et al.,

2022).

As mutações na posição R346 surgiram de forma tardia, R346K e R346T foram prevalentes em variantes BA.1.* e BA.4.6/BQ.1.1, respectivamente. Resultados experimentais sugerem que essas mutações podem conferir uma resistência à neutralização por anticorpos previamente identificados por atuarem em抗ígenos de SARS-CoV-2 (CAO et al., 2022). A mutação V213G surge de maneira tardia a partir de BA.2.* e a mutação T95I a partir da variante Delta AY.101 e prevalece ao longo das sub-linhagens Ômicron BA.1.*.

A glicoproteína spike é essencial para a interação com a enzima de conversão de angiotensina 2 (*ACE2*, sigla em inglês), expresso nas células hospedeiras, e é importante para a transmissão viral, sendo uma região crítica para prevalência de mutações *missense* (HARVEY et al., 2021; PICCOLI et al., 2020; TOYOSHIMA et al., 2020). Na Figura 9B e 9C, é possível observar que o domínio de ligação de receptor (*Receptor Binding Domain - RBD*) acumula o maior número de mutações em relação ao C-terminal e N-terminal, região de maior contato com receptores *ACE2*.

De acordo com Piccoli et al. (2020), o RBD de SARS-CoV-2 é imunodominante em relação a quantidade total de anticorpos monoclonais elicitados, é alvo de 90% da atividade neutralizante presente no soro ou plasma de indivíduos infectados e apresenta dois sítios principais na neutralização da fusão do vírus à células hospedeiras pela ligação ao *ACE2*. Além da região *RBD*, as variantes Ômicron possuíram mutações conservadas (H655Y, N679K e P681H) adjacente ao *FCS*, as quais potencializaram a clivagem da proteína *spike* e a fusão com a célula hospedeira (VIANA et al., 2022).

No entanto, ao contrário das variantes Alpha (P681H) e Delta (P681R), as mutações em *RBD* e *FCS* não foram determinantes para prevalência e alta transmissibilidade das variantes Ômicron. As mudanças genotípicas nas novas variantes do vírus demonstraram alterar o fenótipo viral, modulando as respostas imunes inatas e a evasão da resposta imune adaptativa, além de modificar a funcionalidade da proteína *spike*, afetando a transmissão e a patogênese. (HARVEY et al., 2021; WILLETT et al., 2022). Segundo resultados apresentados por Willett et al. (2022), variantes Ômicron BA.1 e BA.2 mudaram a preferência de via de entrada à célula hospedeira, utilizando um mecanismo de fusão endossômica independente de TMPRSS2, principal protease transmembrana envolvida na fusão do envelope viral.

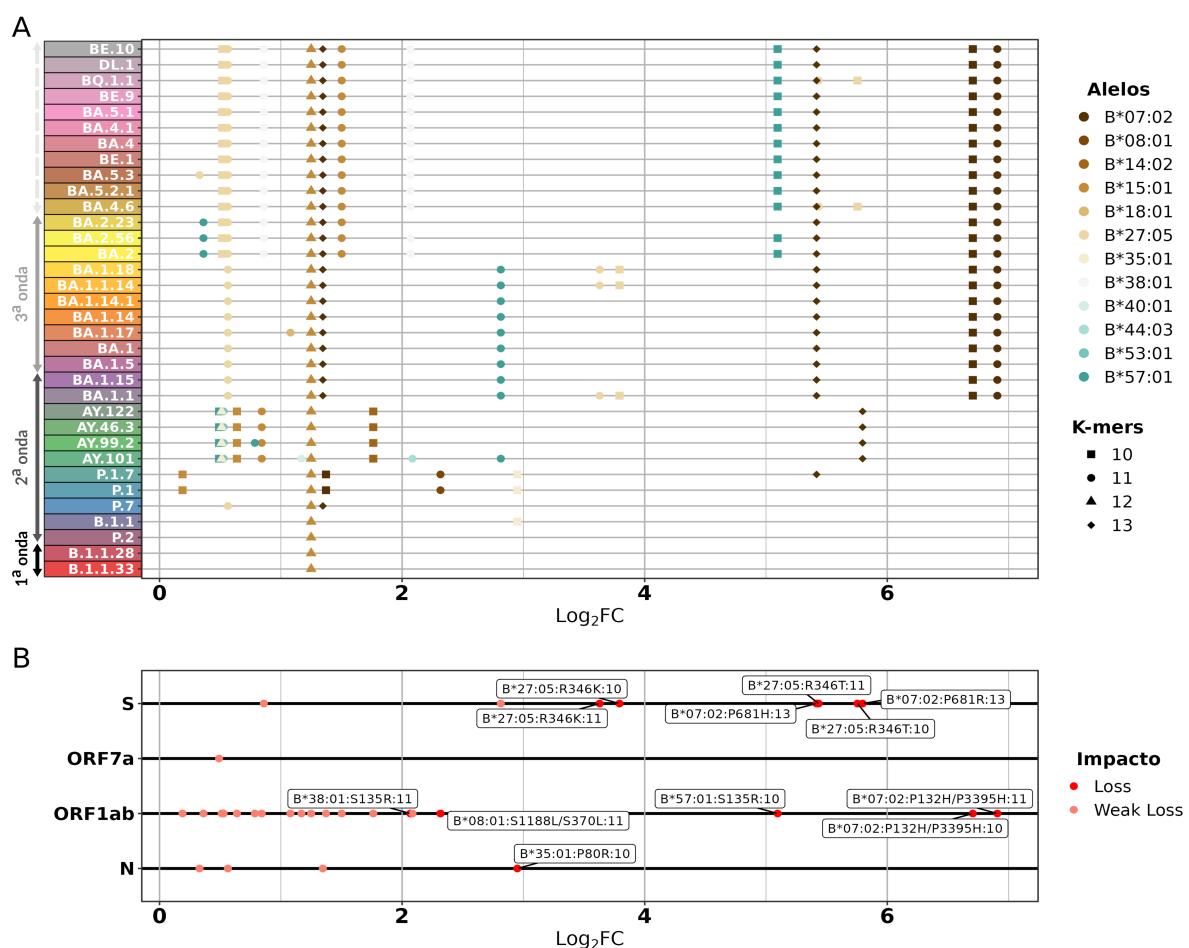
Em Foz do Iguaçu, além da D614G, P681* e H655Y que prevalecem na *spike* de diferentes variantes, as mutações mais abundantes ocorreram ao longo da circulação de sub-linhagens Ômicron, evidenciando a diversidade de mutações dessas linhagens que geraram uma série de mecanismos fenotípicos adaptados para sua prevalência ao longo do tempo. Apesar da mutação D614G surgir amplamente nas variantes da primeira onda, a segunda mutação mais frequente H655Y é compartilhada exclusivamente entre as variantes Gamma P.1 e Ômicron BA.1.1 (OU et al., 2022).

5.3 Epítópos de B*07:02 apresentam maior perda na afinidade de ligação com mutações *missense*

A partir do *pipeline* desenvolvido, foram detectadas 23 epítópos mutados que apresentaram impacto negativo nos valores de afinidade de ligação preditos em relação aos alelos escolhidos de HLA-B. De acordo com a Figura 10A, as variantes apresentaram um acúmulo de mutações classificadas em Weak Loss/Loss ao longo do tempo. O surgimento de mutações com perdas leves preditas teve maior concentração em ORF1ab (Figura 10A e 10B) e quando ocorreram efeitos negativos, mutações na proteína *spike* geraram um alto impacto na afinidade de ligação predita.

Figura 10 - Valores de $\text{Log}_2(\text{Fold Change})$ para cada mutação classificada em (Weak) Loss.

Mutações classificadas por (A) variantes e (B) por genes. Variantes estão ordenadas de baixo para cima a partir da primeira data de detecção em Foz do Iguaçu. Mutações que geraram Loss estão identificadas com alelo:mutação:*k-mer* e indicação na lateral esquerda da onda epidemiológica que ocorreu a primeira detecção em (B)



Fonte: O Autor (2024)

Tabela 2 - Mutações *Loss* epítópos associados ao alelo B*07:02 e B*27:05

Epítopo (Ref/mut)	K-mers	Gene	Alelo	Mutação	$\log_2(FC)$	FC^1	Núm. de variantes
RP/HNFTIKGSFL	11	ORF1ab	B*07:02	P132H/P3395H	6.9	120.03	23
RP/HNFTIKGSF	10	ORF1ab	B*07:02	P132H/P3395H	6.7	104.34	23
SP/RRAARSVASQSI	13	S	B*07:02	P681R	5.79	55.57	4
ATR/TFASVYAW	10	S	B*27:05	R346T	5.75	53.98	2
TR/TFASVYAWNR	11	S	B*27:05	R346T	5.43	43.23	2
SP/HRRARSVASQSI	13	S	B*07:02	P681H	5.41	42.72	24
ATR/KFASVYAW	10	S	B*27:05	R346K	3.79	13.85	3
TR/KFASVYAWNR	11	S	B*27:05	R346K	3.62	12.36	3
IPARARVE/DCF	10	ORF1ab	B*07:02	E341D	1.37	2.58	2
GPQNQRNAP/LRITF	13	N	B*07:02	P13L	1.34	2.54	24
QNQRNAP/LRITF	11	N	B*27:05	P13L	0.56	1.47	24
ARLR/CAKHVVY	10	ORF1ab	B*27:05	R5716C/R392C	0.51	1.43	14
ARLR/CAKHVVY	10	ORF1ab	B*27:05	R5716C/R392C	0.51	1.43	14
RRIRGGDGKM/IK	11	N	B*27:05	M101I	0.32	1.25	1

¹Fold Change

Fonte: O Autor (2024)

Em relação ao perfil de mutações *Weak Loss/Loss* de cada variante, é possível visualizar que foi diferente em cada período marcado pelas ondas epidemiológicas. Como apresentado na Figura 9A, variantes Gamma, Zeta e Delta marcam essa mudança, seguida pelas variantes Ômicron. Curiosamente, as mutações nas primeiras variantes detectadas e com menor magnitude de *fold change* concentram-se em grande parte na *ORF1ab*. Essa região codifica proteínas não estruturais envolvidas na replicação viral e podem tolerar mais mutações, pois estão envolvidas em processos que podem ser menos impactados por mutações pontuais.

Os epítópos mutados classificados com perda apresentaram 1.1-*fold* à 120-*fold* de diminuição nos valores preditos em relação à sequências de referência. Enquanto isso, os epítópos que apresentaram maior amplitude ($\log_2(Fold\ Change) \geq 5$), entre 34-*fold* à 120-*fold*), possuem prevalência a partir das variantes Gamma (P.1.7) e permanecem ao longo das variantes Ômicron. As perdas são detectadas quando comparadas a epítópos de referência B*07:02, seguido dos epítópos de B*27:05 (Figura 10B). O momento de surgimento das mutações que atingem epítópos B*07:02 difere ao período de maior número de pacientes internados que carregam o grupo alélico B*07 ao final da primeira onda, segundo os dados experimentais.

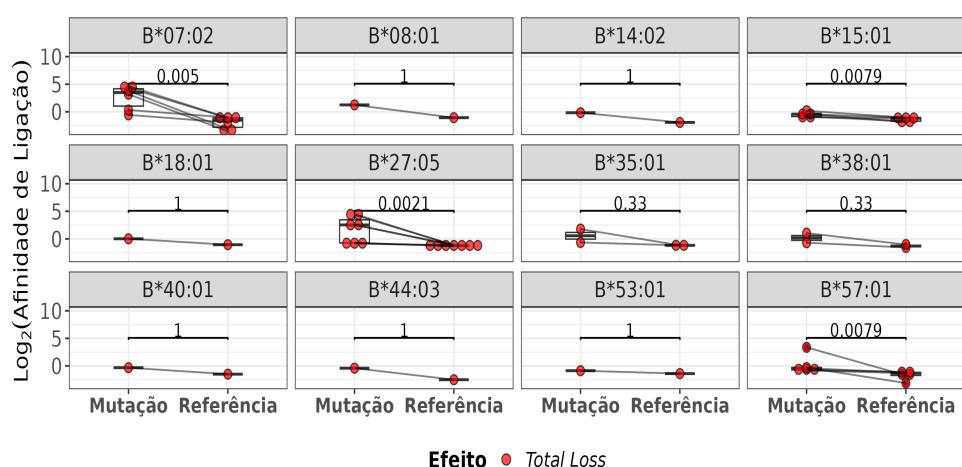
A mutação Prolina (P) → X e Arginina (R) → X afetam, majoritariamente, os valores preditos para epítópos B*07:02 e B*27:05, respectivamente. Especialmente, quando localizado na posição âncora do epítopo, nesse caso, P2. Conforme visualizado na (Tabela 2), esse tipo de mutação está diretamente relacionada com o aumento da magnitude de *fold change*. Dados similares foram apresentados em estudos anteriores, onde coletaram dados genômicos globais de SARS-CoV-2 e verificaram o impacto negativo nesse tipo de mutação na apresentação alélica de epítópos associados ao HLA-B7 *supertype*, mas um impacto positivo foi apontado para HLA-B27 *supertype* (HAMELIN et al., 2022).

Em relação a alta prevalência, os epítópos gerados pelas mutações P323L/P4715L

(VLFSTVFPP/LTSF) do *RdRp* (*ORF1ab*) estão presentes no genoma de todas as variantes que circularam em Foz do Iguaçu, sendo a mutação mais prevalente fora da região da proteína *spike* e foi classificada como *Weak Loss* em relação aos epítopos B*15:01. Estudos anteriores apontam que a mutação P323L favorece a estabilidade estrutural da *nsp12* e afeta diretamente a sua função por meio da interação da sua posição com co-fatores *nsp7* e *nsp8* - aumentando a taxa de mutação no processo de tradução (KIM et al., 2023; KIRCHDOERFER; WARD, 2019). Além disso, observa-se uma coevolução significativa entre as mutações D614G e P323L, na qual a presença isolada de G614 ou L323 não ganhou relevância epidemiológica. Em contrapartida, a combinação dessas duas mutações resultou em uma variante viral G/L que quase completamente substituiu a variante original D/P (ILMJÄRV et al., 2021).

Por fim, é possível identificar que epítopos associados ao B*07:02 e B*27:05 obtiveram uma maior magnitude de *fold change* a partir do surgimento de mutações (Figura 11). Apesar de dados *in silico* não confirmarem o papel das mutações no escape imunogênico de epítopos associados à B*27:05 como apontado para B*07:02, estudos experimentais com células CD8+ corroboram com o impacto negativo de mutações em epítopos associados a ambos alelos (WELLINGTON et al., 2023).

Figura 11 - Comparação entre a predição da afinidade de ligação em escala logarítmica (eixo y) de epítopos com mutações e referência (eixo x) para cada alelo. *Total Loss* são todas as classificações *Weak Loss* e *Loss* e p-valores se encontram na parte superior de cada quadro.



Fonte: O Autor (2024)

5.4 Implicações e limitações do estudo na região de fronteira

Ao final de 2020, o surgimento de Variantes de Interesse (*VOI*, sigla em inglês) e Variantes de Preocupação (*VOC*, sigla em inglês) gerou a ocorrência de novas ondas epidemiológicas, apresentando perfil de transmissibilidade e patogênese diferente quando comparada às variantes iniciais. A origem de mutações nas novas variantes com alta prevalência instigaram buscar compreender seus impactos no escape do sistema imune humano e considerá-las fatores importantes durante o desenvolvimento de vacinas ou guiando na decisão de políticas públicas no combate ao vírus.

O município de Foz do Iguaçu está localizado em uma área de tríplice fronteira com potencial exportação e importação de novas variantes, ou seja, servindo como área estratégica na contenção de ondas epidemiológicas causadas por novas variantes. Situada na fronteira entre Brasil, Paraguai e Argentina, a cidade possui fronteiras terrestres entre os três países e com uma atividade socioeconômica baseada no turismo e no trânsito comercial aduaneiro. Durante a pandemia de COVID-19, a cidade contou com a flexibilização de medidas de segurança sanitária, especialmente, a permanência do funcionamento de hotéis, aeroporto, rodoviária e a reabertura das fronteiras (RIVAS et al., 2020).

Os laboratórios de detecção molecular de SARS-CoV-2 foram empregados para realizar um monitoramento de casos no município, localizados no Hospital Municipal Padre Germano Lauck (HMPGL) e Hospital Ministro Costa Cavalcanti (HMCC). Os esforços na detecção de casos de COVID-19 também se estenderam no rastreamento de variantes, realizando coleta e sequenciamento de genomas de SARS-CoV-2 em diferentes períodos epidemiológicos. No início da pandemia, adicionalmente, a Universidade Federal da Integração Latino-Americana (UNILA) apresentou um amplo estudo buscando compreender a imunidade humoral e celular de casos assintomáticos no município (VIANA et al., 2021).

Apesar desses dados serem armazenados em bancos públicos, como o GISAID e DATASUS, ou publicados em artigos científicos em conjunto com dados de demais regiões do Brasil, não houve uma abordagem que examinasse a circulação de variantes especificamente no município (GIOVANETTI et al., 2022). Por isso, o presente trabalho buscou relacionar estudos experimentais prévios sobre o perfil alélico de HLA-B de pacientes admitidos na UTI e o perfil genômico de SARS-CoV-2 no mesmo período, empregando ferramentas computacionais para destacar o panorama de mutações que prevaleceram na cidade e que potencialmente puderam causar um impacto negativo no sistema de apresentação de antígenos humano.

Esse trabalho conseguiu caracterizar as ondas epidemiológicas, podendo descrever a prevalência das variantes ao longo do tempo e suas respectivas mutações. Algo importante para a vigilância sanitária na região de fronteira, visto que estudos mais robustos identificaram a exportação de novas variantes para o Paraguai (GIOVANETTI et al., 2022). Quando comparado com o estudo conduzido por Giovanetti et al. (2022),

é possível identificar que a Figura 7 apresenta maior similaridade na distribuição dos períodos de surgimento e co-circulação das variantes Zeta e Gamma entre Foz do Iguaçu e Paraguai, em comparação com a distribuição apresentada para o Brasil.

Na Figura 7, nota-se ainda que houve um maior número de pacientes que possuem o grupo alélico B*07 admitidos na UTI e que vieram a óbito ao final da primeira onda, período que havia maior presença das variantes B.1.1.28 e B.1.1.33. Em contrapartida, essas variantes não apresentaram mutações que impactaram os epítopes de HLA-B*07:02 como reportado para variantes Ômicron, conforme apresentado na Figura 9.

Estudos anteriores destacaram a reatividade cruzada de células T B*07:02+, observando que indivíduos não infectados, mas portadores do alelo HLA-B*07:02, conseguem reconhecer o peptídeo N(105–113), e extremamente conservado, derivado do SARS-CoV-2 devido à presença de células T reativas cruzadas que reconhecem o peptídeo homólogo N(105–113) dos coronavírus OC43-CoV e HKU1-CoV. Além de estar associado com uma menor progressão da infecção e maior conservação da eficiência antiviral pós-infecção (FRANCIS et al., 2021; PENG et al., 2022).

Essa discordância com a literatura pode estar relacionada com as limitações do estudo experimental, onde a genotipagem a partir do sequenciamento foi realizada manualmente, com baixa resolução e não realizou-se técnicas estatísticas robustas para estratificação das amostras com base na idade, vacinação (para pacientes de 2021), estilo de vida dos pacientes e prevalência dos grupos alélidos na população local. Esse último pode justificar a alta presença de pacientes portadores de B*07 e nenhum portador de B*27, os quais apresentam alta e baixa frequência populacional, respectivamente (Apêndice A).

Por outro lado, a análise *in silico* evidenciou o impacto negativo das mutações nas vias dependentes de HLA-B*07:02 e B*27:05, em concordância com estudos experimentais prévios (WELLINGTON et al., 2023). Adicionalmente, reproduziu resultados sobre o impacto de posições de mutações determinantes na eficiência da apresentação de抗ígenos, conforme apontado nos valores preditos para P→X e R→X em epítopes B*07:02 e B*27:05, respectivamente (HAMELIN et al., 2022).

Esses dados *in silico* podem direcionar tecnologias de vacinação ou até mesmo, quando relatados previamente, auxiliar na tomada de decisões de políticas públicas sanitárias quando somado a demais dados epidemiológicos e clínicos da população local. Em um estudo recente, pesquisadores demonstraram que determinados alelos HLA-II são determinantes na resposta humoral de indivíduos a partir da vacinação, mas não é o determinante para predizer a resposta clínica e novos avanços da COVID-19 de forma isolada e outros fatores devem ser considerados, especialmente, os haplótipos de HLA (OLAFSDOTTIR et al., 2022).

Em Foz do Iguaçu, Viana et al. (2022) conduziram um levantamento do perfil de soroconversão e resposta imune celular, a partir de inquéritos sorológicos, com indivíduos assintomáticos ao longo do mês de Maio à Setembro de 2020. Observou-se que, quando

extrapolado à nível de população, é perdido 25% da capacidade de soroconversão anti-SARS-CoV-2 em três meses e uma diminuição significativa em cinco meses para antígenos provenientes da região *RBD*.

Cabe ressaltar que durante os últimos inquéritos já havia relatos da circulação da variante Zeta no Brasil, a qual já carregava mutações na região *spike* e próximas ao RBD (GIOVANETTI et al., 2022). Mesmo não necessitando de dados da dinâmica mutacional do vírus, os resultados contrapuseram a política de imunidade de rebanho defendida pelo governo naquele momento (GURGEL, 2021; UNILA, 2020). Adicionalmente, indicando também que as tecnologias de vacinação possuem potencial risco de serem afetadas se não houver um olhar atento aos dados genômicos das variantes.

Contudo, o presente estudo *in silico* apresentou limitações computacionais, exigindo a redução do número de alelos escolhidos para as análises e concentrando-se apenas no HLA-B; as predições foram conduzidas com mutações *missense* individuais, sem considerar a soma de mutações ao longo das variantes; a abordagem buscou identificar apenas a potencial perda na afinidade de ligação a partir das mutações, no entanto, o ganho pode ser mais elusivo para o desenvolvimento de novas tecnologias; e foram utilizadas as classificações das variantes disponíveis no GISAID, mas classificações filogenéticas *ad hoc* podem ser empregados para avaliar de melhor forma a dinâmica populacional do vírus na cidade, assim como, obter maior acurácia na descoberta de novas mutações.

6 CONCLUSÃO

O presente trabalho apresentou o repertório das variantes de SARS-CoV-2 que circularam no município de Foz do Iguaçu-Paraná e o panorama do impacto das mutações em alelos HLA-B prevalentes na população local. Esse fluxo de análises fornece um método básico com possibilidade de aprimoramento e que pode ser empregado no estudo de outros agentes infecciosos. Os presentes resultados podem auxiliar futuros estudos na área de vigilância genômica na região e contribuir para os aspectos da medicina personalizada, oferecendo dados para o desenvolvimento de imunizantes com base no perfil genômico da população local.

REFERÊNCIAS

- ARANTES, I. et al. Emergence and Spread of the SARS-CoV-2 Variant of Concern Delta across Different Brazilian Regions. **Microbiology Spectrum**, v. 10, n. 5, p. e02641–21, ago. 2022. Disponível em: <<https://journals.asm.org/doi/10.1128/spectrum.02641-21>>.
- BASTOS, L. S. et al. COVID-19 hospital admissions: Brazil's first and second waves compared. **The Lancet Respiratory Medicine**, v. 9, n. 8, p. e82–e83, ago. 2021. ISSN 2213-2600, 2213-2619. Disponível em: <[https://www.thelancet.com/journals/lanres/article/PIIS2213-2600\(21\)00287-3/fulltext](https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(21)00287-3/fulltext)>.
- BENTON, D. J. et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. **Nature**, v. 588, n. 7837, p. 327–330, dez. 2020. ISSN 1476-4687. Disponível em: <<https://www.nature.com/articles/s41586-020-2772-0>>.
- CAO, Y. et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. **Nature**, v. 602, n. 7898, p. 657–663, fev. 2022. ISSN 1476-4687. Disponível em: <<https://www.nature.com/articles/s41586-021-04385-3>>.
- CASTRO, M. d. S. et al. High-resolution characterization of 12 classical and non-classical HLA loci in Southern Brazilians. **HLA**, v. 93, n. 2-3, p. 80–88, fev. 2019. ISSN 2059-2310.
- FARIA, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. **Science**, v. 372, n. 6544, p. 815–821, maio 2021. Disponível em: <<https://www.science.org/doi/full/10.1126/science.abh2644>>.
- FRANCIS, J. M. et al. Allelic variation in class I HLA determines CD8+ T cell repertoire shape and cross-reactive memory responses to SARS-CoV-2. **Science Immunology**, v. 7, n. 67, p. eabk3070, nov. 2021. Disponível em: <<https://www.science.org/doi/10.1126/sciimmunol.abk3070>>.
- GFELLER, D. et al. Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8+ T-cell epitopes. **Cell Systems**, v. 14, n. 1, p. 72–83.e5, jan. 2023. ISSN 2405-4712, 2405-4720. Disponível em: <[https://www.cell.com/cell-systems/abstract/S2405-4712\(22\)00470-7](https://www.cell.com/cell-systems/abstract/S2405-4712(22)00470-7)>.
- GIOVANETTI, M. et al. Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. **Nature Microbiology**, v. 7, n. 9, p. 1490–1500, set. 2022. ISSN 2058-5276. Disponível em: <<https://www.nature.com/articles/s41564-022-01191-z>>.
- GONZALEZ-GALARZA, F. et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. **Nucleic Acids Research**, v. 48, n. D1, p. D783–D788, jan. 2020. ISSN 0305-1048. Disponível em: <<https://doi.org/10.1093/nar/gkz1029>>.
- GRIFONI, A. et al. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. **Cell Host & Microbe**, v. 27, n. 4, p. 671–680.e2, abr. 2020. ISSN 19313128. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1931312820301669>>.

- GRÄF, T. et al. Dispersion patterns of SARS-CoV-2 variants Gamma, Lambda and Mu in Latin America and the Caribbean. **Nature Communications**, v. 15, n. 1, p. 1837, fev. 2024. ISSN 2041-1723. Disponível em: <<https://www.nature.com/articles/s41467-024-46143-9>>.
- GURGEL, B. **Imunidade de rebanho e atraso de vacinas impulsionaram mortes, diz relatório**. 2021. Disponível em: <<https://www.cnnbrasil.com.br/politica/imunidade-de-rebanho-e-atraso-de-vacinas-impulsionaram-mortes-diz-relatorio/>>.
- HAMELIN, D. J. et al. The mutational landscape of SARS-CoV-2 variants diversifies T cell targets in an HLA-supertype-dependent manner. **Cell Systems**, v. 13, n. 2, p. 143–157.e3, fev. 2022. ISSN 24054712. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405471221003811>>.
- HARRISON, P. W. et al. Ensembl 2024. **Nucleic Acids Research**, v. 52, n. D1, p. D891–D899, jan. 2024. ISSN 0305-1048. Disponível em: <<https://doi.org/10.1093/nar/gkad1049>>.
- HARVEY, W. T. et al. SARS-CoV-2 variants, spike mutations and immune escape. **Nature Reviews Microbiology**, v. 19, n. 7, p. 409–424, jul. 2021. ISSN 1740-1526, 1740-1534. Disponível em: <<http://www.nature.com/articles/s41579-021-00573-0>>.
- HOOF, I. et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. **Immunogenetics**, v. 61, n. 1, p. 1–13, jan. 2009. ISSN 1432-1211. Disponível em: <<https://doi.org/10.1007/s00251-008-0341-z>>.
- ILMJÄRV, S. et al. Concurrent mutations in RNA-dependent RNA polymerase and spike protein emerged as the epidemiologically most successful SARS-CoV-2 variant. **Scientific Reports**, v. 11, n. 1, p. 13705, jul. 2021. ISSN 2045-2322. Disponível em: <<https://www.nature.com/articles/s41598-021-91662-w>>.
- KHARE, S. et al. GISAID’s Role in Pandemic Response. **China CDC Weekly**, v. 3, n. 49, p. 1049–1051, 2021. ISSN 2096-7071. Disponível em: <<http://weekly.chinacdc.cn/en/article/doi/10.46234/ccdw2021.255>>.
- KIM, S.-M. et al. SARS-CoV-2 variants with NSP12 P323L/G671S mutations display enhanced virus replication in ferret upper airways and higher transmissibility. **Cell Reports**, v. 42, n. 9, p. 113077, set. 2023. ISSN 2211-1247. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2211124723010884>>.
- KINSELLA, R. J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. **Database**, v. 2011, p. bar030, jan. 2011. ISSN 1758-0463. Disponível em: <<https://doi.org/10.1093/database/bar030>>.
- KIRCHDOERFER, R. N.; WARD, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. **Nature Communications**, v. 10, n. 1, p. 2342, maio 2019. ISSN 2041-1723. Disponível em: <<https://www.nature.com/articles/s41467-019-10280-3>>.
- KIYOTANI, K. et al. Bioinformatic prediction of potential T cell epitopes for SARS-CoV-2. **Journal of Human Genetics**, v. 65, n. 7, p. 569–575, jul. 2020. ISSN 1434-5161, 1435-232X. Disponível em: <<http://www.nature.com/articles/s10038-020-0771-5>>.

- KLEIN, J.; SATO, A. The HLA System. **New England Journal of Medicine**, v. 343, n. 10, p. 702–709, set. 2000. ISSN 0028-4793. Disponível em: <<https://doi.org/10.1056/NEJM200009073431006>>.
- 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. 812–827.e19, ago. 2020. ISSN 0092-8674, 1097-4172. Disponível em: <[https://www.cell.com/cell/abstract/S0092-8674\(20\)30820-5](https://www.cell.com/cell/abstract/S0092-8674(20)30820-5)>.
- LISTA, M. J. et al. The P681H Mutation in the Spike Glycoprotein of the Alpha Variant of SARS-CoV-2 Escapes IFITM Restriction and Is Necessary for Type I Interferon Resistance. **Journal of Virology**, v. 96, n. 23, p. e01250–22, nov. 2022. Disponível em: <<https://journals.asm.org/doi/full/10.1128/jvi.01250-22>>.
- MOHR, C. et al. **nf-core/epitopeprediction: v2.3.1 - Oesterberg - 2024-05-17**. Zenodo, 2024. Disponível em: <<https://zenodo.org/records/11209196>>.
- MOURA, E. C. et al. Covid-19: temporal evolution and immunization in the three epidemiological waves, Brazil, 2020–2022. **Revista de Saúde Pública**, v. 56, p. 105, dez. 2022. ISSN 0034-8910, 1518-8787. Disponível em: <<https://www.scielo.br/j/rsp/a/798jKxCNGhB85QBJXdK6h9z/?lang=en>>.
- OLAFSDOTTIR, T. A. et al. HLA alleles, disease severity, and age associate with T-cell responses following infection with SARS-CoV-2. **Communications Biology**, v. 5, n. 1, p. 1–14, set. 2022. ISSN 2399-3642. Disponível em: <<https://www.nature.com/articles/s42003-022-03893-w>>.
- OU, J. et al. Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. **Signal Transduction and Targeted Therapy**, v. 7, n. 1, p. 1–9, abr. 2022. ISSN 2059-3635. Disponível em: <<https://www.nature.com/articles/s41392-022-00992-2>>.
- O'DONNELL, T. J.; RUBINSTEYN, A.; LASERSON, U. MHCflury 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. **Cell Systems**, v. 11, n. 1, p. 42–48.e7, jul. 2020. ISSN 2405-4712, 2405-4720. Disponível em: <[https://www.cell.com/cell-systems/abstract/S2405-4712\(20\)30239-8](https://www.cell.com/cell-systems/abstract/S2405-4712(20)30239-8)>.
- PENG, Y. et al. An immunodominant NP105–113-B*07:02 cytotoxic T cell response controls viral replication and is associated with less severe COVID-19 disease. **Nature Immunology**, v. 23, n. 1, p. 50–61, jan. 2022. ISSN 1529-2916. Publisher: Nature Publishing Group. Disponível em: <<https://www.nature.com/articles/s41590-021-01084-z>>.
- PEREZ, M. A. S. et al. Structural Prediction of Peptide–MHC Binding Modes. In: SIMONSON, T. (Ed.). **Computational Peptide Science: Methods and Protocols**. New York, NY: Springer US, 2022. p. 245–282. ISBN 978-1-07-161855-4. Disponível em: <https://doi.org/10.1007/978-1-0716-1855-4_13>.
- PETERS, B.; NIELSEN, M.; SETTE, A. T Cell Epitope Predictions. **Annual Review of Immunology**, v. 38, p. 123–145, abr. 2020. ISSN 1545-3278.

- PICCOLI, L. et al. Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. **Cell**, v. 183, n. 4, p. 1024–1042.e21, nov. 2020. ISSN 0092-8674, 1097-4172. Disponível em: <[https://www.cell.com/cell/abstract/S0092-8674\(20\)31234-4](https://www.cell.com/cell/abstract/S0092-8674(20)31234-4)>.
- PONTAROTTI, P.; PAGANINI, J. COVID-19 Pandemic: Escape of Pathogenic Variants and MHC Evolution. **International Journal of Molecular Sciences**, v. 23, n. 5, p. 2665, fev. 2022. ISSN 1422-0067. Disponível em: <<https://www.mdpi.com/1422-0067/23/5/2665>>.
- RAMMENSEE, H. et al. SYFPEITHI: database for MHC ligands and peptide motifs. **Immunogenetics**, v. 50, n. 3-4, p. 213–219, nov. 1999. ISSN 0093-7711.
- REYNISSON, B. et al. NetMHCpan-4.1 and NetMHCIIPan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. **Nucleic Acids Research**, v. 48, n. W1, p. W449–W454, jul. 2020. ISSN 0305-1048, 1362-4962. Disponível em: <<https://academic.oup.com/nar/article/48/W1/W449/5837056>>.
- RIVAS, A. V. et al. Increase SARS-CoV-2 infection in a border region after Brazil national holidays / Aumento da infecção de SARS-CoV-2 em uma região de fronteira após feriados nacionais no Brasil. **Brazilian Journal of Development**, v. 6, n. 12, p. 101252–101259, dez. 2020. ISSN 2525-8761. Disponível em: <<https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/22055>>.
- SHAO, X. M. et al. High-Throughput Prediction of MHC Class I and II Neoantigens with MHCnuggets. **Cancer Immunology Research**, v. 8, n. 3, p. 396–408, mar. 2020. ISSN 2326-6066. Disponível em: <<https://doi.org/10.1158/2326-6066.CIR-19-0464>>.
- SIEGRIST, C.-A. 2 - Vaccine Immunology. In: PLOTKIN, S. A. et al. (Ed.). **Plotkin's Vaccines (Seventh Edition)**. Elsevier, 2018. p. 16–34.e7. ISBN 978-0-323-35761-6. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B978032335761600002X>>.
- TAY, M. Z. et al. The trinity of COVID-19: immunity, inflammation and intervention. **Nature Reviews Immunology**, v. 20, n. 6, p. 363–374, jun. 2020. ISSN 1474-1733, 1474-1741. Disponível em: <<http://www.nature.com/articles/s41577-020-0311-8>>.
- TOYOSHIMA, Y. et al. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. **Journal of Human Genetics**, v. 65, n. 12, p. 1075–1082, dez. 2020. ISSN 1434-5161, 1435-232X. Disponível em: <<http://www.nature.com/articles/s10038-020-0808-9>>.
- UNILA. GT da UNILA estuda dinâmicas da pandemia em diferentes cenários de isolamento social.** 2020. Disponível em: <<https://portal.unila.edu.br/noticias/gt-da-unila-estuda-dinamicas-da-pandemia-em-diferentes-cenarios-de-isolamento-social-1>>.
- VIANA, K. et al. Infection in asymptomatic carriers of SARS-CoV-2 can interfere with the achievement of robust immunity on a population scale. **Journal of General Virology**, v. 102, n. 11, p. 001684, 2021. ISSN 1465-2099. Disponível em: <<https://www.microbiologyresearch.org/content/journal/jgv/10.1099/jgv.0.001684>>.

- VIANA, R. et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. **Nature**, v. 603, n. 7902, p. 679–686, mar. 2022. ISSN 1476-4687. Disponível em: <<https://www.nature.com/articles/s41586-022-04411-y>>.
- V'KOVSKI, P. et al. Coronavirus biology and replication: implications for SARS-CoV-2. **Nature Reviews Microbiology**, v. 19, n. 3, p. 155–170, mar. 2021. ISSN 1740-1526, 1740-1534. Disponível em: <<http://www.nature.com/articles/s41579-020-00468-6>>.
- WELLINGTON, D. et al. SARS-CoV-2 mutations affect antigen processing by the proteasome to alter CD8+ T cell responses. **Helix**, v. 9, n. 10, out. 2023. ISSN 2405-8440. Disponível em: <[https://www.cell.com/helix/abstract/S2405-8440\(23\)07284-5](https://www.cell.com/helix/abstract/S2405-8440(23)07284-5)>.
- WIECZOREK, M. et al. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. **Frontiers in Immunology**, v. 8, mar. 2017. ISSN 1664-3224. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fimmu.2017.00292/full>>.
- WILLETT, B. J. et al. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. **Nature Microbiology**, v. 7, n. 8, p. 1161–1179, ago. 2022. ISSN 2058-5276. Disponível em: <<https://www.nature.com/articles/s41564-022-01143-7>>.
- WIRTH, W.; DUCHENE, S. **GISAIDR**. Zenodo, 2022. Disponível em: <<https://zenodo.org/records/6474693>>.
- XIA, B. et al. Why is the SARS-CoV-2 Omicron variant milder? **The Innovation**, v. 3, n. 4, jul. 2022. ISSN 2666-6758. Disponível em: <[https://www.cell.com/the-innovation/abstract/S2666-6758\(22\)00047-9](https://www.cell.com/the-innovation/abstract/S2666-6758(22)00047-9)>.
- YANG, H.; RAO, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. **Nature Reviews Microbiology**, v. 19, n. 11, p. 685–700, nov. 2021. ISSN 1740-1526, 1740-1534. Disponível em: <<https://www.nature.com/articles/s41579-021-00630-8>>.

APÊNDICES

APÊNDICE A – Frequência Relativa HLA-B

A.1 Proporção relativa dos vinte alelos HLA-B mais frequentes na região do Brasil, Estado do Paraná e Curitiba/PR.

	<i>Brazil Mixed AFND</i> ¹		Estado do Paraná (REDOME) ²		<i>Curitiba - Paraná</i> ³	
	Alelos	Proporção Relativa	Alelos	Proporção Relativa	Alelos	Proporção Relativa
1	B*07:02:01	0.087	B*35	0.121	B*07:02	0.082
2	B*18:01	0.051	B*44	0.107	B*44:03	0.077
3	B*35:01	0.051	B*15	0.088	B*15:01	0.067
4	B*51:01:01	0.051	B*51	0.086	B*51:01	0.058
5	B*08:01	0.043	B*07	0.072	B*18:01	0.053
6	B*08:05	0.036	B*08	0.056	B*35:01	0.053
7	B*44:03	0.036	B*18	0.053	B*38:01	0.038
8	B*39:01	0.029	B*14	0.05	B*08:01	0.034
9	B*44:02	0.029	B*40	0.049	B*49:01	0.034
10	B*08:03	0.022	B*39	0.036	B*14:02	0.029
11	B*14:02	0.022	B*57	0.027	B*35:03	0.029
12	B*15:01:01	0.022	B*27	0.026	B*39:01	0.029
13	B*35:02	0.022	B*38	0.026	B*44:02	0.029
14	B*35:05	0.022	B*49	0.026	B*27:05	0.024
15	B*38:01:01	0.022	B*52	0.022	B*35:02	0.024
16	B*42:02	0.022	B*50	0.021	B*40:01	0.024
17	B*53:01	0.022	B*13	0.019	B*53:01	0.024
18	B*58:01	0.022	B*58	0.019	B*57:01	0.024
19	B*78:01	0.022	B*53	0.017	B*13:02	0.019
20	B*07:02:03	0.014	B*45	0.014	B*15:17	0.019

¹ Gonzalez-Galarza et al. (2020)

² <https://redome.inca.gov.br/>

³ Castro et al. (2019)