Title: SARS-CoV-2 evolution in the absence of selective immune pressures, results in antibody resistance, interferon suppression and phenotypic differences by lineage.

Authors: Julian Daniel Sunday Willett[1,2,3], Annie Gravel[4], Isabelle Dubuc[4], Leslie Gudimard[4], Ana Claudia dos Santos Pereira Andrade[4], Émile Lacasse[4], Paul Fortin[4,5,6], Ju-Ling Liu[2,8], Jose Avila Cervantes[2,8], Jose Hector Galvez[7], Haig Hugo Vrej Djambazian[2,8], Melissa Zwaig[2,8], Anne-Marie Roy[2,8], Sally Lee[2,8], Shu-Huang Chen[2,8], Jiannis Ragoussis[2,8]*, Louis Flamand[4,9]*

Affiliations:

1. Quantitative Life Sciences Ph.D. Program, McGill University, Montreal, QC, Canada
2. McGill Genome Centre, McGill University, Montreal, QC, Canada
3. Lady Davis Institute, Jewish General Hospital, Montreal, QC, Canada
4. Axe maladies infectieuses et immunitaires, Centre de Recherche du Centre Hospitalier Universitaire de Québec- Université Laval, Canada
5. Centre de Recherche ARThrite-Arthrite, Recherche et Traitements, Université Laval, Québec, QC, Canada
6. Division of Rheumatology, Department of Medicine, CHU de Québec-Université Laval, Québec, QC, Canada
7. Canadian Centre for Computational Genomics, McGill University, Montreal, QC, Canada
8. Department of Human Genetics, McGill University, Montreal, QC, Canada
9. Département de microbiologie-infectiologie et d'immunologie, Université Laval, QC, Canada

*Co-corresponding authors

Abstract:

The persistence of COVID-19 is partly due to viral evolution reducing vaccine and treatment efficacy. Serial infections of Wuhan-like SARS-CoV-2 in Balb/c mice yielded mouse-adapted strains with greater infectivity and mortality. We investigated if passaging unmodified B.1.351 (Beta) and B.1.617.2 (Delta) 20 times in K18-ACE2 mice, expressing human ACE2 receptor, in a BSL-3 laboratory without selective pressures, would drive human health-relevant evolution and if evolution was lineage-dependent. Late-passage virus caused more severe disease, at organism and lung tissue scales, with late-passage Delta demonstrating antibody resistance and interferon suppression. This resistance co-occurred with a *de novo* spike S371F mutation, linked with both traits. S371F, an Omicron-characteristic mutation, was co-inherited at times with spike E1182G per Nanopore sequencing, existing in different quasi-species at others. Both are linked to mammalian GOLGA7 and ZDHHC5 interactions, which mediate viral-cell entry and antiviral response. This study demonstrates SARS-CoV-2's tendency to evolve with phenotypic consequences, its evolution varying by lineage, and suggests non-dominant quasi-species contribute.

Keywords: SARS-CoV-2, COVID-19, pandemic modelling, treatment resistance, viral evolution.

Introduction:

Part of the difficulty in responding to the COVID-19 pandemic has been predicting viral evolution and its impact on clinically relevant traits, such as disease severity, infectivity, and treatment resistance[1]. The basic biology of the severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2), the virus that causes COVID-19, is well understood[2]. However, knowledge of SARS-CoV-2 evolution is limited. Many RNA viruses, like SARS-CoV-2, exist as quasi-species meaning that within a given viral population, a multitude of alleles (major and minor) are present. These alleles give plasticity to the viral population allowing for rapid adaptation/selection to a variety of circumstances[3]. Viral evolution is also complex and affected by several factors, including host genetic background[4], host immune status[5], the organs targeted by the virus[6], and a population's collective immunity[7]. As a result, it has been difficult to predict the emergence of new variants with altered virulence.

Several SARS-CoV-2 variants of concern (VOC) have emerged, each with mutations providing dissemination advantages[1]. The SARS-CoV-2 Alpha variant (B.1.1.7) gained a growth advantage and rapidly spread globally due to the spike protein N501Y mutation that enhanced affinity for the cellular entry receptor, angiotensin-converting enzyme 2 (ACE2)[8]. Other VOC, such as the Delta (B.1.617.2) variant, gained advantages with additional spike substitutions, including L452R[9], T478K[10], and P681R[11] that affect viral transmissibility and antibody neutralization for naturally or artificially vaccinated individuals[12]. To better predict changes in SARS-CoV-2 and reduce the impact of a future pandemic, time-effective models must be defined for studying viral changes over successive generations and characterizing their evolutionary consequences.

In 2020, Gu *et al.* studied short-term viral evolution by infecting aged mice with the SARS-CoV-2 reference strain (IME-BJ05) and collecting lung tissue at set timepoints following infection, using

this isolate to infect successive generations of mice[13]. While wild-type mice are considered to be less susceptible to the virus since mouse Ace2 is not used by SARS-CoV-2 as its receptor, Gu *et al.* demonstrated that serial passaging of infected lung homogenate across mice produced a mouse-adapted strain (MASCp6) that causes pulmonary symptoms, matching adaptations observed by others who previously applied the technique to influenza[13,14]. After six passages, they found several clinically relevant genetic changes, including the *de novo* appearance of A23063T that causes the spike mutation N501Y[13]. This mutation is associated with enhanced viral infection and transmission[8]. Subsequently, the group passaged the MASCp6 isolate an additional 30 times and characterized the MASCp36 isolate[15]. MASCp36 contained several mutations (K417N, Q493H, N501Y) in the spike protein conferring greater mouse Ace2 receptor binding affinity, increased infectivity and greater virulence[15]. Others replicated this study using the same viral variant, observing similar development of clinically relevant variant alleles[16].

Since these studies, new VOC have arisen with the latest being Omicron, negatively impacting treatment efficacy[1]. Others have studied viral dynamics using recent lineages in serial passages in cells, observing similar development of directional selection and clinically concerning mutations with contributions by quasi-species[17,18]. Understanding longitudinal viral changes in a mammalian host is pressing. We adapted Gu *et al.*'s approach[13] to determine if more recent and human-adapted lineages, B.1.351 (Beta) and B.1.617.2 (Delta), evolved over a twenty passage study in transgenic mice expressing the human ACE2 receptor, if this evolution varied by lineage, and whether these changes were relevant to human health. We used transgenic animals to limit selective pressure and adaptations to the murine ACE2 receptor, as occurred in Gu *et al.*'s works[13,15]. We report on virus evolving in the setting of no experimental selective pressures with changes observed at the organism, tissue, and genetic scale, this evolution varying by lineage with solely late-study Delta lineage presenting characteristics resulting in

decreased antibody neutralization, with variant alleles arising *de novo* that have been linked with these phenotypes.

Results:

Late passaged virus produced more viral RNA, greater weight loss, and worse lung inflammation. Only late-study Delta was associated with increased viral load and cytokine changes linked to severe COVID-19. To study evolution on the organism level, we compared clinical scores of mice infected with early and late passaged virus (**Figure 1**). Unlike Gu *et al*.'s study that required adaptation of SARS-CoV-2 to a murine Ace2 receptor for infectivity, our study used transgenic mice expressing the human ACE2 receptor[13]. On day 3 post-infection, lungs of passage [P]20 Beta and Delta infected mice contained significantly more SARS-CoV-2 RNA than corresponding P0 viruses (**Figure 2A**). Mice infected with P20 Delta virus had a significant increase (6.3x) in lung viral loads relative to P0 (**Figure 2B**) (p=0.008) with no difference for Beta P0 and P20 (**Figure 2B**). Both P20 Beta and Delta viruses caused more rapid and greater weight loss than P0 (**Figures 2C-D**) with similar survival curves (data not shown).

To determine if the significant increase in lung viral load observed by Delta virus was related to growth kinetic differences between P0 and P20 viruses, we performed *in vitro* growth curve analysis. Using two independently evolved P20 Delta viruses, we observed that P20 viruses yielded significantly (4-5X) more infectious virus after 24h of infection versus Delta P0 virus (p = 0.02) (**Figure 3**). This difference in viral titers was no longer observed after 48 h of infection (**Figure 3**).

Lungs of mice infected with P0 and P20 viruses were examined for signs of inflammation as assessed by histology and presence of inflammatory cytokines. On day 6 post-infection, P20

Beta and Delta viruses caused greater leukocyte infiltration, red blood cell extravasation, and decreased alveolar space versus P0 and mock-infected mice (**Figures 4A-E**). Lung tissues were also analyzed for SARS-CoV-2 antigenic burden and leukocyte recruitment as measured using anti-N and anti-CD45 antibodies, respectively. Results indicated that significantly more cells were expressing N by Beta P20 than P0, with no change by passage for Delta (**Figure 5**). While Delta had significantly greater immune cell infiltration than Beta per CD45 staining, immune infiltration did not significantly change between P0 and P20 for either lineage (**Figure 5**). On day 3 post infection, Beta P20 virus induced greater production of proinflammatory CXCL1, CCL2 and IL-6 versus P0 (**Figures 6A-C**)[19]. Delta P20 infection resulted in diminished IFN-β and IFN-γ production versus P0, which correlated with the increased viral load and has been linked with more severe COVID-19 in humans[20,21] (**Figures 6D-E**).

We next assessed whether repeated passages in K18-ACE2 mice would select for viruses capable of infecting non transgenic mice. C57BL6 (B6) mice were infected intranasally with P0 and P20 Beta and Delta viruses and lung SARS-CoV-2 RNA and infectious viral loads determined 3 days later. Both P0 and P20 Beta viruses successfully replicated in B6 mouse lungs with both viruses producing similar SARS-CoV-2 RNA and infectious viral loads (**Supplementary Figure 1**). None of the P0 or the P20 Delta virus infected mice had detectable SARS-CoV-2 RNA or infectious viral loads above the limit of detection (**Supplementary Figure 1**).

Late-study Delta, not Beta, virus presented with greater antibody resistance. We next determined the ability of sera from vaccinated subjects to neutralize passaged viruses, comparing it to stock P0 viruses. Consistent with previous reports, sera from vaccinated subjects had varying neutralizing potential depending on the variant, with neutralization titers being greatest against the original Wuhan-like strain and lowest for the Beta isolate (**Figure**

7A)[22,23]. The same sera were analyzed for their ability to neutralize P20 viruses. P20 Beta was neutralized equally (13 out of 24) or more efficiently (8 of 24) that the P0 virus (**Figure 7B**). In contrast, P20 Delta virus was significantly less sensitive to antibody neutralization by vaccinated subject sera than P0 ($p < 0.0002$) with 15 of 24 sera showing decreased neutralizing potential toward the P20 Delta virus relative to P0 (**Figure 7C**).

Variant alleles associated with COVID-19 severity and susceptibility, cytokine suppression, and antibody resistance arose de novo predominantly in late passaged Delta virus. To link phenotypic changes with genetics, we identified those variant alleles that changed in frequency across passages, including those that arose *de novo* and disappeared, and annotated them for traits with implications on public health, such as conferring antibody resistance. Variant allele frequency across passages and between samples in single passages did not statistically change (adjusted p-value ≥ 0.05), except when comparing Beta P17 and P20 to earlier passages and Delta P10 and P20 (**Supplemental Figure 2a-b**). Given no significant change in mean allele depth, these differences could be due to growing allelic diversity and within-host evolution of quasi-species, per Tonkin-Hill *et al*.[24]. Sample contamination was less likely as there was no significant difference in VAF or variant allele depth between samples for either viral lineage (**Supplemental Figure 2c-d**).

Of 82 unique variant alleles, there were 26 that were nonsynonymous and appeared *de novo* or whose allele frequency (VAF) clearly changed with 20 having immediately clinically-relevant annotations, with 20 occurring in viral proteins that interact with human proteins, 1 believed to confer drug resistance, 1 associated with the Omicron lineage, 1 linked with antibody escape, and 7 at vaccine targets (**Table 1, Supplemental Table 1**) [25,26]. These included 9 that were specific to Beta, 16 that were specific to Delta, and 1 that changed in both lineages. The variant allele associated with Omicron was C22674T (Spike S371F) that arose *de novo* in 2/3 animals

in P20 for Beta virus and 2/3 animals in P13 for Delta virus, persisting in later passages (**Figure 8, Supplemental Table 1**). We thus observed an Omicron specific mutation in a study not involving Omicron viral lineages. The S371F mutation alters spike glycoprotein that interacts with human proteins GOLGA7 and ZDHHC5, which contribute to viral-cell entry[27] and palmitoylation that mediates viral infectivity[28], over-response to pathogens, and interferon production[29]. It has also been reported to undermine antibody response.[30]

Of the variant alleles that changed in frequency, many involved viral proteins that interact with human proteins suggested to influence COVID-19 severity and susceptibility, representing druggable targets, and could explain phenotypes observed in this study (**Supplemental Table 1**). A929G (I222V), which arose and persisted in 1/3 animals infected with Delta in P17 (**Supplemental Figure 3**), involves nsp2 via ORF1ab that interacts with GIGYF2[31]. C10341T and C10809T occur in nsp5 via ORF1ab (P3359L and P3515L), that interacts with HDAC2 that has been identified to enable immune evasion as an anti-immune effector[32]. C10341T arose in 1/3 animals in P13 for Delta and likely persisted for remaining passages, with its absence in P17 perhaps due to sequencing error (**Supplemental Figure 3**). C10809T disappeared after P0 in Beta, reappearing in 1/3 animals in P17 (**Supplemental Figure 3**). HDAC2 has four approved gene family-specific inhibitors, such as vorinostat, and one approved activator, theophylline[33]. G28237T disappeared after P0, then reappeared in 3/3 animals infected with Beta by P20 (**Figure 8**). It occurred in ORF8 (R115L), which interacts with LOX, linked to COVID-19 severity and thrombosis[34], PLOD2, linked to COVID-19 and respiratory failure[35], and FKBP10, linked to poor COVID-19 prognoses[36]. *LOX* can be targeted by 3 approved medications, including the inducer cupric sulfate[33], *PLOD2* can be targeted by three approved medications, including a cofactor ascorbic acid[33], and FKBP10 can be stimulated by bleomycin[37].

Most alleles in the P0 Beta isolate, passaged several times *in vitro* in Vero cells, encoded a Spike protein with a tryptophan (W) at position 682 (C23606T), destroying the furin cleavage

site (**Table 1**). We observed a rapid selection of viruses having the reference allele encoding an arginine (R) at position 682 and functional furin site in all animals infected with Beta variant starting in P10 (**Table 1**). This suggests that the furin site provides a growth advantage under *in vivo* conditions and supports a previous report indicating the Spike furin site being important for SARS-CoV-2 virulence in K18-ACE2 mice[38].

Nanopore focused sequencing of the S (spike) gene suggested several of these variant alleles were co-inherited by distinct Delta quasi-species. Spike mutations S371F and E1182G were observed to occur at identical Nanopore-derived allele frequencies in P17 for animal two and P20 for animal three, suggesting co-inheritance from P13 viruses. While only S371F and A892V were observed to have the same allele frequency in P13 with E1182G not detected by Nanopore in this passage, but detected by Illumina, this could be a product of sequencing error or changes, alongside observing co-inheritance of alleles at distinct times in distinct samples (**Figure 8**). S371F confers antibody escape with S371F and E1182G both interacting with GOLGA7 and ZDHHC5, which enable viral cell entry[27] and infectivity[28].

Discussion:

In this work, we show that more recent and human-adapted lineages of SARS-CoV-2, Beta and Delta, evolve in a setting of minimal selective pressures, Delta developed more clinically relevant changes than Beta with both yielding more pronounced lung disease and disease severity, and these phenotypic changes can be partially explained by discrete development and disappearance of alleles linked with key traits, such as antibody resistance and interferon suppression. Given our observance of these alleles with frequencies suggestive of minor quasi-species, it could suggest their significance in COVID-19, experimentally supporting others' conclusions in a controlled setting[24]. While the evolution of the original SARS-CoV-2 strain in

Balb/c mice has been previously reported, with genetic changes used to explain phenotypic evolution[13,15], Beta and Delta lineages are more genetically adapted to humans with additional spike mutations, making a model more similar to humans with mice expressing human ACE2R more pressing. This also limits selection for murine Ace2 adaptation, establishing a baseline for a model to build off. Furthermore, none have compared the phenotypic changes of two lineages in a single study, offering controlled, experimental evidence for lineage-based evolutionary differences.

The Delta P0 virus, which lacked the N501Y mutation, never acquired it in our study. Acquisition of N501Y in humans was reported almost a year after SARS-CoV-2 epidemic's beginning, suggesting the mutation arose alongside growing herd immunity[39]. This could suggest that in the absence of selective immune pressure, such as neutralizing antibodies, or pressure to adapt to murine ACE2, N501Y is not favored (**Figure 8**). Beta P0 harbored the N501Y mutation, which was previously reported to enable reference SARS-CoV-2's adaptation in non transgenic mice[13]. This mutation remained unchanged in our study, suggesting that our virus could infect cells through the murine ACE2 receptor, which we demonstrated in non transgenic C57BL6 mice using Beta P0 and P20 viruses (**Supplementary Figure 1**). ACE2 transgene being more widely expressed than murine ACE2 could explain the greater SARS-CoV-2 RNA loads in K18 mouse lungs versus B6. Beta's flexibility to use human and murine receptors could explain our isolate's greater virulence versus others, such as Wuhan and Delta that lacked N501Y[40].

Evolved Beta and Delta modulated the inflammatory and antiviral responses differently than corresponding P0 viruses, both causing greater and more rapid weight loss and more severe lung damage without a significant change in detected inflammatory cells, suggesting mediation by overactivation of existing cells or systemic factors (**Figure 5**). Beta P20 infection was associated with greater inflammatory cytokine production (CCL2, CXCL1, IL-6) than P0 ($p < 0.005$, $p = 0.051$, $p < 0.05$, respectively) (**Figure 6**). This occurred with the disappearance of

several variant alleles, and the *de novo* development of S371F in P20 for 2/3 animals (**Figure 8, Supplemental Figure 3**). G23593T, or spike Q677H, which disappeared from Beta virus samples after P0, enhances treatment resistance and increase viral infectivity [41]. The reversion of Spike amino acid 677 to Q after passage in mice could explain the increased neutralization activity for some of the sera against Beta P20 relative to P0 (**Figure 7**). Delta P20 infection yielded greater viral load than P0 with suppressed IFN-β and IFN-γ, suggesting acquisition of additional antiviral/immunomodulatory properties that could be mediated by S371F's (C22674T) effect on spike protein, which can suppress type I interferon expression[42]. S371F, which arose *de novo* and persisted in two animals in P20 for Beta and P13 for Delta virus (**Figure 8**), is one of the 8 Omicron BA.2-specific spike mutations that induces a 27-fold reduction in the capacity of sotrovimab to neutralize BA.2, broadly affecting the binding of most RBD-directed antibodies (**Table 1**)[30]. A929G (ORF1ab I222V) also arose *de novo* in 1/3 animals, which could impact nsp2's suppression of GIGYF2's functions (**Table 1, Supplementary Figure 3**)[31].

This study had limitations. We used K18-ACE2 transgenic mice but the mechanisms of new viral variants arising in humans could significantly differ, which is relevant as host factors are believed to contribute to viral variant allele diversity[43]. Our model lacked experimental selective pressures, while the COVID-19 pandemic has been accompanied by contact limitations and vaccines with the dominant SARS-CoV-2 evolutionary mechanism believed to be natural selection[7]. A baseline is required to effectively model these pressures. We studied Beta and Delta lineage virus, which are no longer the dominant lineages and our conclusions could vary with Omicron. This underscores the value in studying single-allele changes as they can capture fitness-related traits.

Conclusions:

We demonstrate that more human-adapted SARS-CoV-2 lineages when passaged in mice expressing human ACE2 receptor evolve in the setting of minimal selective pressures, their changes vary by lineage, and accumulate clinically-relevant changes, such as antibody resistance.

Methods:

Viruses. SARS-CoV-2 Wuhan-like strain (LSPQ, B1 lineage) was obtained from the Laboratoire de Santé Publique du Québec ([LSPQ] Sainte-Anne-de-Bellevue, QC, Canada). SARS-CoV-2 Beta strain was obtained from BEI resources and SARS-CoV-2 Delta strain from the BC Centers for Disease Control. Unmodified SARS-CoV-2 strains were propagated on Vero cells (American Type Culture Collection, Manassas, Virginia, USA).

***Determination of the viral titer*.** Vero cells were plated in a 96-well plate ($2x10^4$/well) and infected with 200µl of serial dilution of the viral preparation or lung homogenate in the M199 media supplemented with 10mM HEPES pH 7.2, 1mM of sodium pyruvate, 2.5g/L of glucose, 5µg/mL *Plasmocin*® and 2% FBS. Three days post-infection plates were analyzed using a EVOS M5000 microscope (ThermoFisher Scientific, Waltham, MA, USA) and the viral titer was determined using the Karber method [44].

Mouse models. B6.Cg-Tg(K18-hACE2)2Prlmn/J (stock#3034860) and B6 mice were purchased from the Jackson Laboratories (Bar Harbor, ME). All mouse studies were conducted in a BSL-3 laboratory. One K18-ACE2 mouse was intranasally infected with either B.1.351 (Beta) or B.1.617.2 (Delta) SARS-CoV-2 lineages. Following 3-4 days, the animals were sacrificed and lung homogenate collected. The homogenate was used to intranasally infect sequential animals, defined as passage. This was repeated for 10 passages. After passage 10

(P10), virus in lung homogenate was sequenced. K18-ACE2 mice were then intranasally infected with Beta P10 or Delta P10 with viruses passaged in three mice at a time for an additional 10 times (total of 20 passages) with virus in lung homogenate sequenced following passages 13, 17, and 20 (**Figure 1**). Viral stocks were then made from lung homogenates of Beta and Delta P20 viruses. K18-ACE2 mice were infected intranasally with 500 $TCID_{50}$ of P0 or P20 viruses. Lung viral loads were determined on day 3 post-infection with weight loss and survival monitored daily for up to 9 days.

Infectivity of P0 and P20 viruses to non-transgenic mice. C57BL6 mice were infected intranasally with 3000 $TCID_{50}$ of P0 and P20 Beta and Delta viruses. Three days later, mice were euthanized, and lungs collected for viral load (titer) determination, as per previously describe[45].

RNA extraction: Up to 30 mg of lung tissue were used for RNA extraction using the Bead Mill Tissue RNA Purification Kit and the Bead Mill Homogenizer (Kennesaw, GA), following manufacturer's instructions.

***Droplet digital PCR (ddPCR) quantitation of SARS CoV-2 RNA*.** SARS-CoV-2 viral RNA loads were determined using Droplet Digital PCR (ddPCR) supermix for probes without dUTP (Bio-Rad Laboratories Ltd.) and the QX200 Droplet Digital PCR System Workflow (Bio-Rad Laboratories Ltd.). The ddPCR primers and probes were previously reported[46].

Multiplex cytokines quantification. Cytokines in mouse lung homogenates were measured using a custom ProcartaPlexTM Mouse Mix & Match Panels kit (Invitrogen, Waltham, MA, USA) on the Bio-Plex 200 (Bio-Rad Laboratories Ltd.).

Histological analysis. For each group analyzed (N=5 mice/group), the right lung lobe was extracted, fixed in formalin and paraffin embedded as described[45]. Lung sections were stained with Carstairs staining for histological analysis. Prior to conduct immunofluorescence assay,

lung sections were deparaffined, hydrated then heat-induced epitope retrieval 16h at 60°C with Diva Decloaker solution (Biocare Medical, Pacheco, CA, USA). Immunostainings were preformed to detect SARS-CoV-2 N antigen and leukocyte infiltration using 20µg/mL rabbit anti-N (Rockland chemicals, Limerick, PA, USA)/anti-rabbit IgG Alexa 488 (Jackson Immuno Research lab, West Grove, PA, USA) and 10µg/mL biotinylated anti-CD45 antibodies (BD Bioscience, Franklin Lakes, NJ, USA)/anti Rat IgG Alexa Plus 647 (Thermo Fisher Scientific, Waltham, MA, USA). Slide were imaged using Axioscan 7 instrument (Carl Zeiss Microscopy, New York, USA) then black and white adjustment were performed with Zen lite 3.7 (Carl Zeiss Microscopy). Quantification of positive area signal for N and CD45 staining were performed using Fiji (ImageJ) threshold analyse tools.

Illumina viral sequencing. RNA extracts were processed for reverse transcription (step a) and targeted SARS-CoV-2 amplification using the ARTIC V3 or V4.1 primer scheme (https://github.com/artic-network/primer-schemes/tree/master/nCoV-2019) (step b). Samples were purified (step b) and Nextera DNA Flex library preparation (step c) was performed for Illumina PE150 paired-end amplicon sequencing on a NovaSeq or MiSeq instrument at the McGill Genome Centre using best practices. Each sample was sequenced twice on different days to obtain a target of minimum 10 million reads per sample. The detailed protocols can be accessed with the following links.

Step a: dx.doi.org/10.17504/protocols.io.bjgekjte.

Step b: dx.doi.org/10.17504/protocols.io.ewov18e4ygr2/v2.

Step c: dx.doi.org/10.17504/protocols.io.bjgnkjve.

Nanopore viral sequencing. RNA extracts were processed by reverse transcription with Lunascript[47]. Targeted SARS-COV-2 amplification was performed using five of the 29 Midnight primers[48]. We targeted the 4167 to 5359 bp region with the primer pairs SARSCoV_1200_5. To

target the Spike region we used primers pairs SARSCoV_1200_23, SARSCoV_1200_25 SARSCoV_1200_22 and SARSCoV_1200_24. Amplification of the primers was performed following Freed and Silander protocol[49]. Nanopore library preparation was made following Reiling *et al.* 2020 and libraries were sequenced on the PromethION 24 sequencer with PromethION Flow Cells V.9.4.1 for a total of 10 million reads[50].

***Genome data processing*.** Following sequencing, the reads from each sample were used to call variants using Freebayes v1.3.6 [51] and the results were saved as a VCF file which was used to compare genomic variation across passages. Others have called variants with other tools such as DeepSNV [24] and VarScan [52]. DeepSNV was not used as it only investigates SNVs [53], when indels have been found to confer selective advantages [54]. Freebayes has been observed to perform slightly superiorly than VarScan when calling variants in wastewater samples [55] and in internal benchmarking (data not shown). The process is described briefly as follows: first, the raw reads from each sample were aligned with BWA MEM v0.7.17 [56] to the reference strain genome sequence (NC_045512.2). The resulting alignments were sorted with duplicate reads flagged. The minimum number of aligned reads for each was logged, and each sample was again aligned and processed, then randomly down sampled to match this minimum number of reads. The minimum number of reads per sample in the first part of the study across passage 0 (P0) and P10 was 8,812,604. The minimum number of reads per sample in the second part of the study across P13, P17, and P20 was 34,253,784. All variants with a quality of less than 20 or a depth below 10 were removed from the analysis. Any variant alleles that overlapped with ARTIC sequencing primers were also removed. Merged VCF files, representing the combination of reads across two sequencing batches for each sample, were used for the paper's main results.

Nanopore reads were processed similarly using freebayes, although run a second time using freebayes with the haplotype-basis-alleles option to use statistical priors to further remove noise

from reads. Given that Nanopore was used to validate Illumina calls and evaluate for co-inheritance, its raw output was not down sampled. Its raw output was aligned using minimap2 v2.24, then processed identically to Illumina reads.

Variant annotation. Every variant allele was inputted into COVID-19 Ensembl variant effect predictor (VEP) (https://covid-19.ensembl.org/index.html) to determine functional consequences and the involved gene and amino acid change [57]. Variants whose VAF changed at least once across passages during the study were logged and input to the UCSC Genome Browser (https://genome.ucsc.edu/), checking for annotations of antibody escape, CD8 escape, vaccine targets, drug resistance, or variants of concern [26].

Determination of spike allele co-inheritance. Alleles were determined to be co-inherited if the nanopore, statistically derived allele frequency was identical for multiple alleles in the same passage.

Antibody neutralization studies. Sera from twenty-four healthy vaccinated (3x) individuals were analyzed for neutralizing activity against Wuhan-like, and Beta and Delta P0 and P20 viruses. Serially diluted sera (in quadruplicate) were incubated with 100 $TCID_{50}$ of virus for 1 h at 37˚C before addition to Vero E6 cells. Three days later, the 96-well plates were observed using an inverted microscope for signs of infection. Neutralization titers were determined and defined as the highest serum dilution preventing infection.

Ethics. The study was conducted in accordance with the Declaration of Helsinki, and mouse protocols were approved by the Comité de protection des animaux de l'Université Laval. Our IRB determined that our work did not constitute gain-of-function research as virus was unmodified and represented circulating variants with the study conducted without experimental selective pressures. Sera obtained from humans were procured following ERB approval (ERB #2022-6204).

Statistics and reproducibility. All statistical analyses were conducted using R/4.2.1. All statistical comparisons between passages were corrected for multiple testing using a Benjamini-Hochberg p-value adjustment. Statistical tests included unpaired and paired t-tests and Mann-Whitney tests, as noted in figures. Reproducibility is described in each method.

Source code. All code used in this study, from sample processing to data analysis, are publicly available on Github at the following link: https://github.com/juliandwillett/SARS_CoV_2_Serial_Passaging_Study.

Data availability. VCF files used to complete computational analyses are available by reasonable request.

Declarations:

Competing interest.

All authors declare no competing interests.

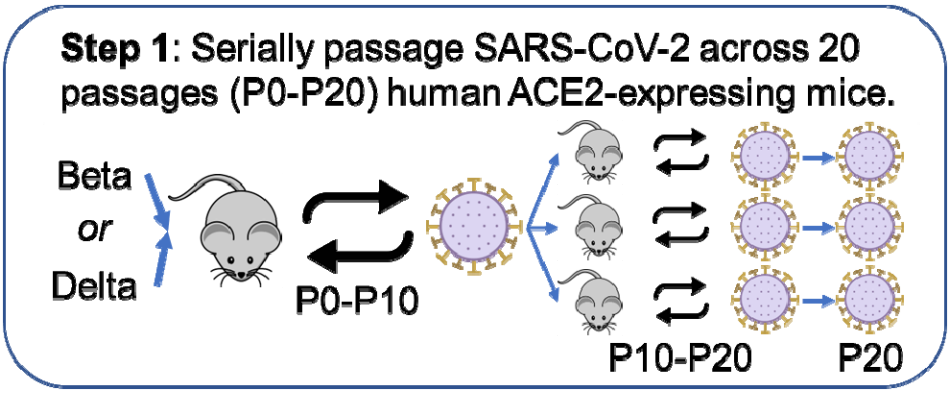
Authors contributions. Julian Daniel Sunday Willett conceived of the computational analyses needed for this study, completed all computational analysis of viral sequencing data, and wrote the majority of the manuscript. Louis Flamand and Jiannis Ragoussis conceived of the study. Louis Flamand contributed to manuscript revisions, wrote sections relevant to animal models and antibody neutralization studies and performed analysis of results. Jiannis Ragoussis supervised sequencing data generation and analysis. Isabelle Dubuc and Leslie Gudimard performed all works involving live SARS-CoV-2 viruses. Annie Gravel participated in the sequence analysis and methodological design of experiments. Ana Claudia dos Santos Pereira Andrade performed the Luminex assay and analyzed the data. Émile Lacasse performed the immunohistological analyses and analyzed the data. Paul Fortin provided key reagents. Jose

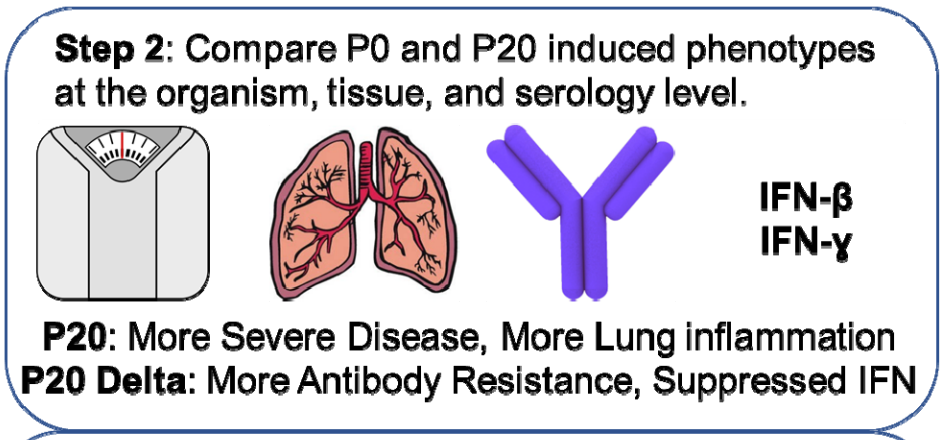
Hector Galvez contributed bioinformatics analysis review and manuscript revisions. Ju-Ling Liu, Jose Avila Cervantes, Haig Hugo Vrej Djambazian, Anne-Marie Roy, Sally Lee and Shu-Huang Chen supported the SARS-CoV-2 viral sequencing.

Acknowledgments.

This research was enabled in part by support provided by Calcul Québec (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca). This study was supported by the Canadian Institutes for Health Research (CIHR 476575) to JDSW, a CIHR operating grant to the Coronavirus Variants Rapid Response Network (CoVaRR-Net) to LF and JR and CIHR grant (GA1-117694) to LF. The funding body had no role in the design of the study or collection, analysis, interpretation, or reporting of the data. We thank Ines Colmegna for providing vaccinated patient sera. We thank Silvia Vidal for her guidance regarding SARS-CoV-2 evolutionary mechanisms. ChatGPT was used to decrease the word count in the abstract with output carefully examined for representing our results. We thank Dr Serge Rivest's imagery platform for the slide scanning service.

Figures





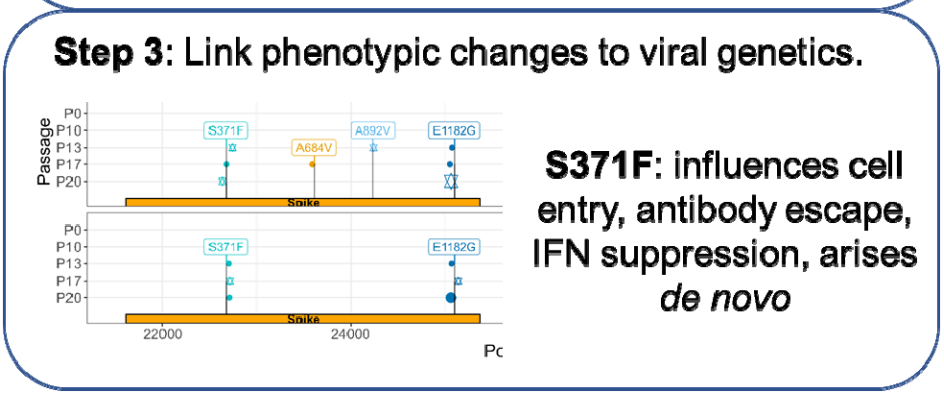


Figure 1. **Study design**. SARS-CoV-2 ACE2-adapted mice were inoculated with either B.1.351 (Beta) or B.1.617.2 (Delta) virus and left for three days for virus to proliferate. They were then sacrificed with the lung homogenate used to infect another mouse. This was repeated for ten passages. P10 virus for each lineage was then used to infect three additional mice and the process was repeated for an additional ten passages. P0 and P20 virus for each lineage was subsequently used to infect additional mice to compare weight loss and survival between early

429 and late passage virus, and test non-adapted mouse susceptibility to the virus. Antibody

430 neutralization of virus was also compared.

431

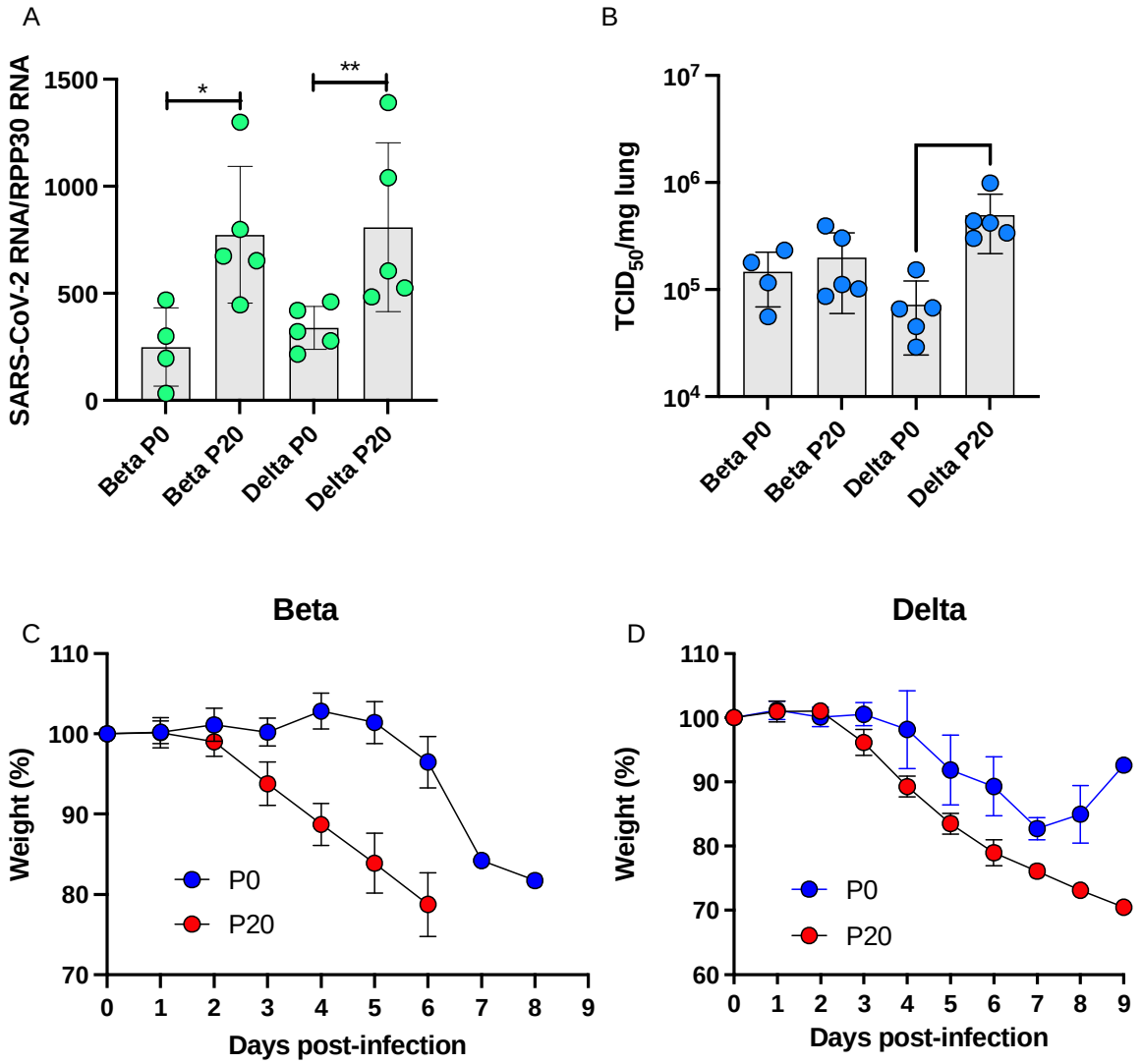
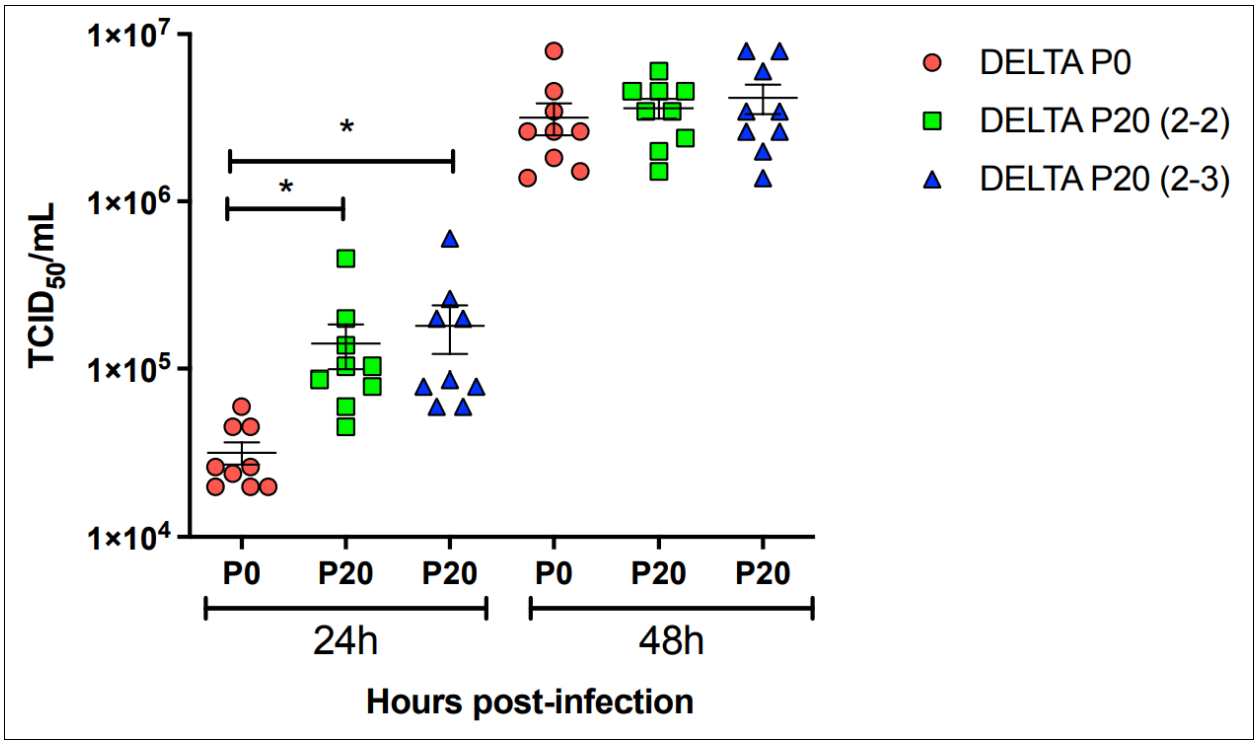


Figure 2. Pathogenesis of P0 and P20 viruses in K18-ACE2 mice. Mice (n = 15/group) were infected with 500 $TCID_{50}$ of P0 or P20 Beta and Delta viruses. (A-B) On day 3 post-infection, mice (n = 5) were euthanized and lungs collected for SARS-CoV-2 RNA (A) and infectious viral load (B) determination. Results are expressed as mean ± SD SARS-CoV-2 RNA copies/RPP30 RNA copies (A) and mean ± SD SARS-CoV-2 $TCID_{50}$/mg of lung tissue (B). (C-D) The weight of each mouse (n = 10/group) was monitored daily throughout the experiment and reported as mean percentage ± SD of weight relative to day 0. * ≤ 0.05 and **: $p \leq 0.008$ as determined using non-parametric Mann-Whitney test.



441 **Figure 3**. **Growth kinetics of P0 and P20 Delta viruses**. P0 and two independently evolved
442 P20 Delta viruses were used to infect Vero cells at a multiplicity of infection of 0.005. Cell-free
443 supernatants were collected at 24 h and 48 h and used for viral titration. Results are expressed
444 as mean ± SD $TCID_{50}$/mL. *: p=0.02 as determined using an unpaired t-test.

445

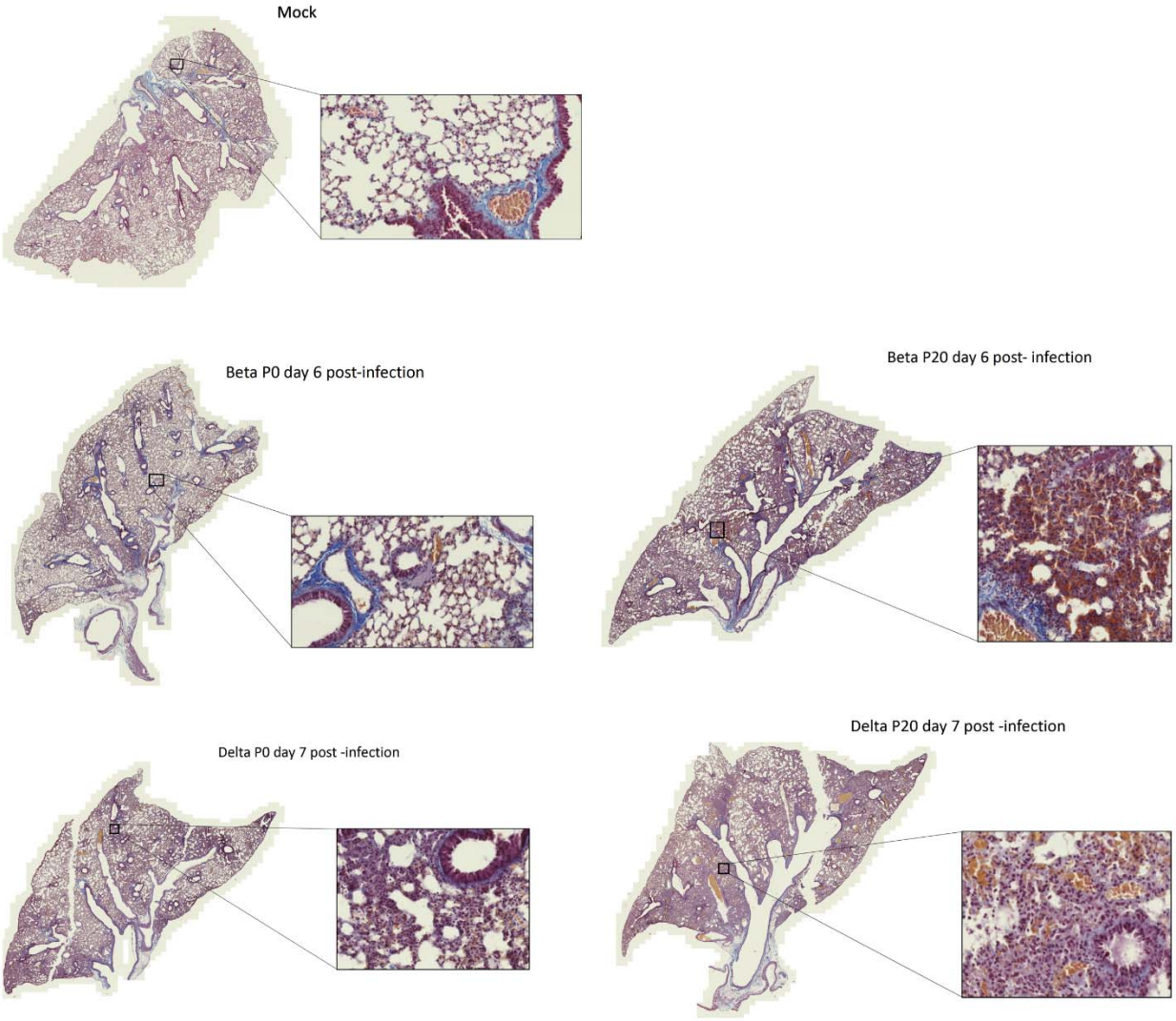
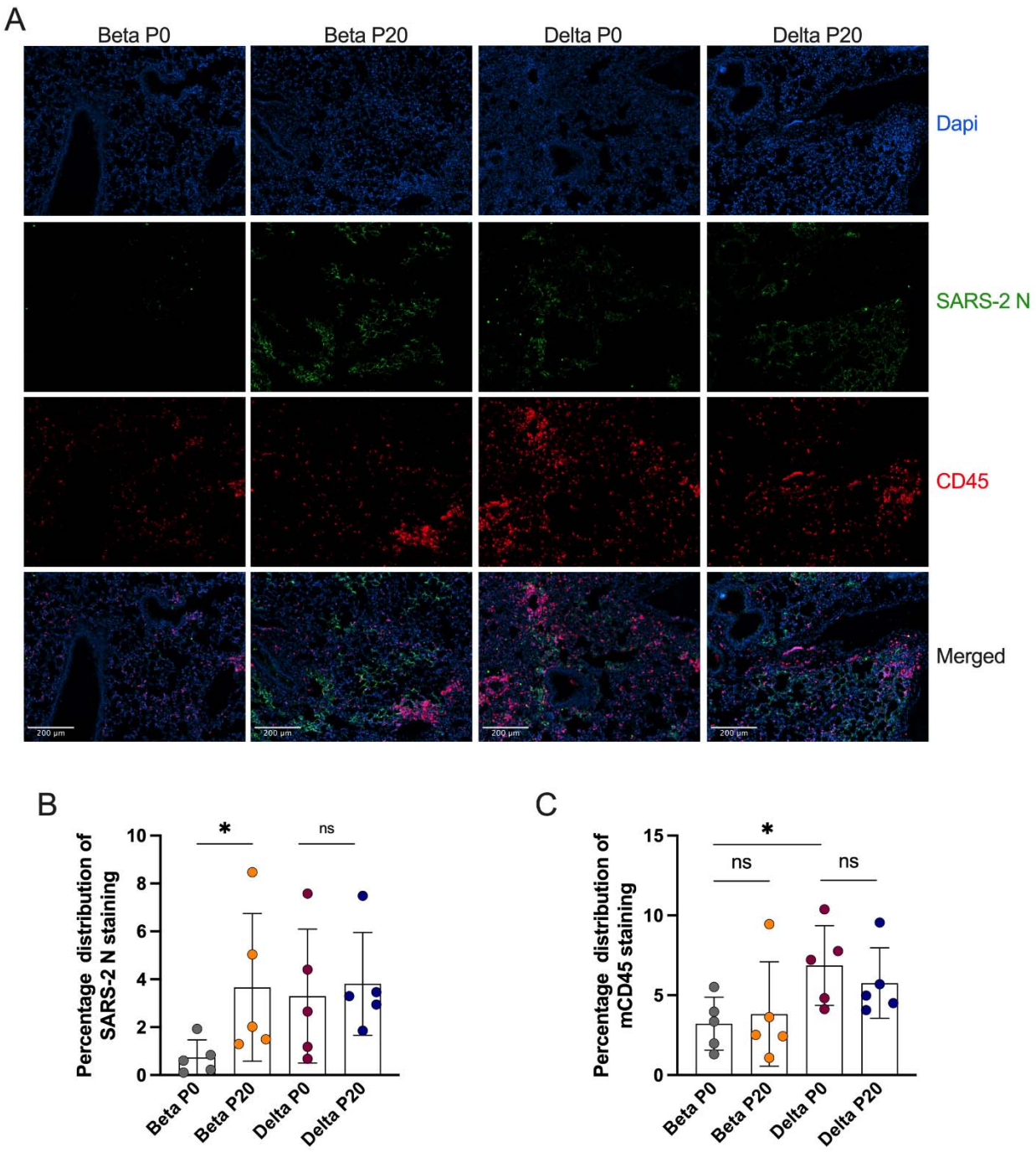


Figure 4. Lung histology of mice infected with Beta and Delta P0 and P20 viruses. Mice were mock-infected (A) or infected with 500 $TCID_{50}$ of Beta P0 (B), Beta P20 (C), Delta P0 (D) of Delta P20 viruses. Lungs were harvested on day 6 (Beta) or day 7 (Delta) post infection, formalin fixed and processed for Carstairs staining.



452

453 **Figure 5**. Immunohistological staining of mouse lungs infected with Beta and Delta P0 and P20

454 viruses. (A) Infected lung tissues were stained with DAPI (blue/nuclei), anti-SARS CoV-2

455 nucleocapsid (N) protein (green), anti-mouse CD45 (red). (B) Graphical (mean +/-SD)

456 representation of SARS CoV-2 N (B) or CD45 (C) staining. Each dot represents data collected

457 from one section of lung tissue from one mouse. Entire lung sections were imaged with

458 representative images shown. *p<0.05 and N.S determined using Mann-Whitney test.

459

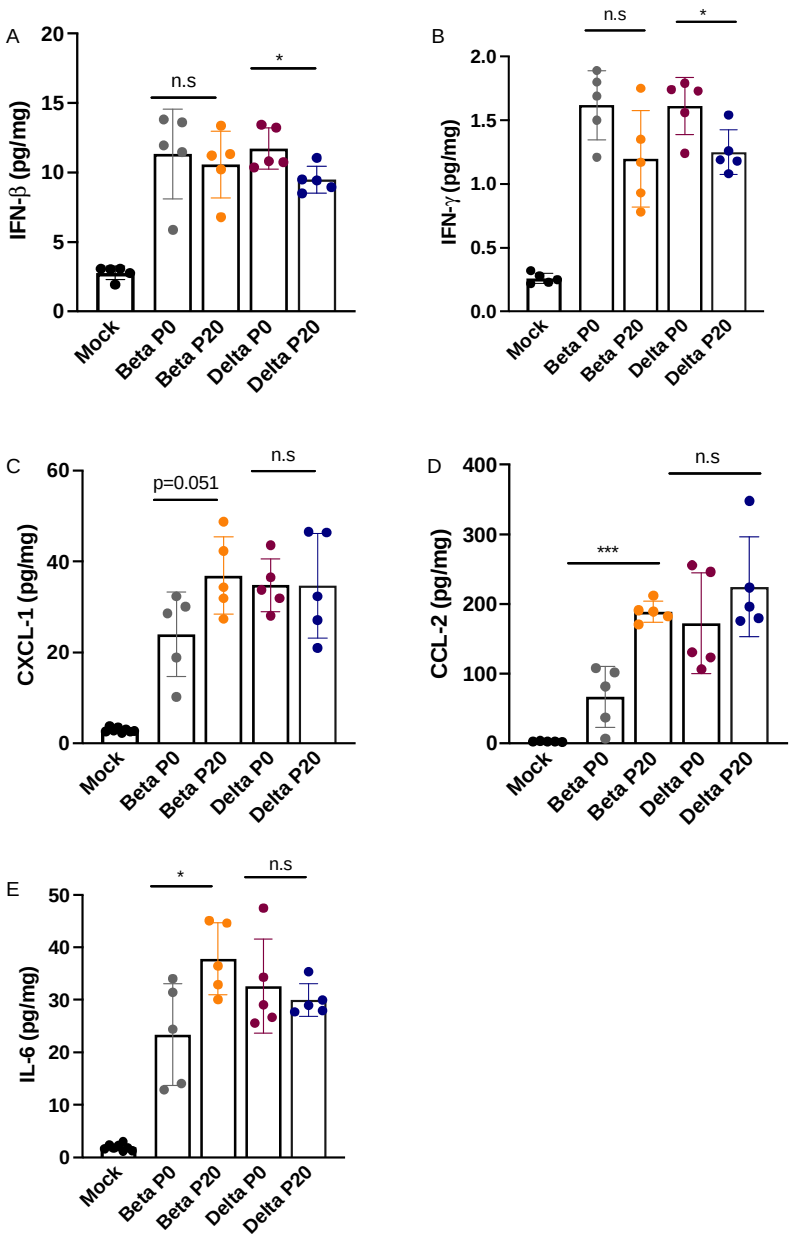


Figure 6. Monitoring of cytokines in lungs of mice infected with Beta and Delta P0 and P20 viruses. K18-ACE2 mice were mock-infected or infected with 500 $TCID_{50}$ of Beta P0, Beta P20 Delta P0 and Delta P20 viruses. On day 3 post-infection, mice were euthanized, lungs were harvested and homogenized. The cytokine content in the homogenates were determines using the Luninex assay. *p<0.05, ***p<0.005 as determined using either unpaired t-test (A, D, E) or Mann-Whitney test (B).

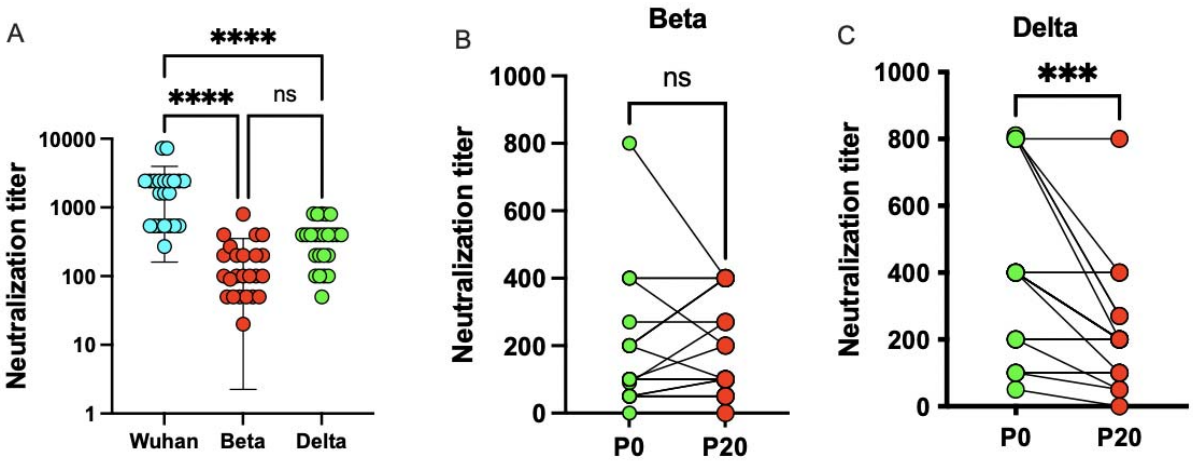


Figure 7. Neutralization of P0 and P20 viruses with sera from vaccinated subjects. (A) Neutralization assay of Wuhan-like, Beta, and Delta viruses using sera from 24 vaccinated subjects. Individual results and mean ± SD neutralization titer against different SARS-CoV-2 isolates are presented. ****: p<0.0001 determined using one way-ANOVA. ns: not statistically significant. (B) Pair-wise comparison of the sera used in A in neutralization assay against P0 and P20 Beta viruses. (C) Pair-wise comparison of the sera used in A in neutralization assay against P0 and P20 Delta viruses. ***: p>0.002 determined using paired t-test.

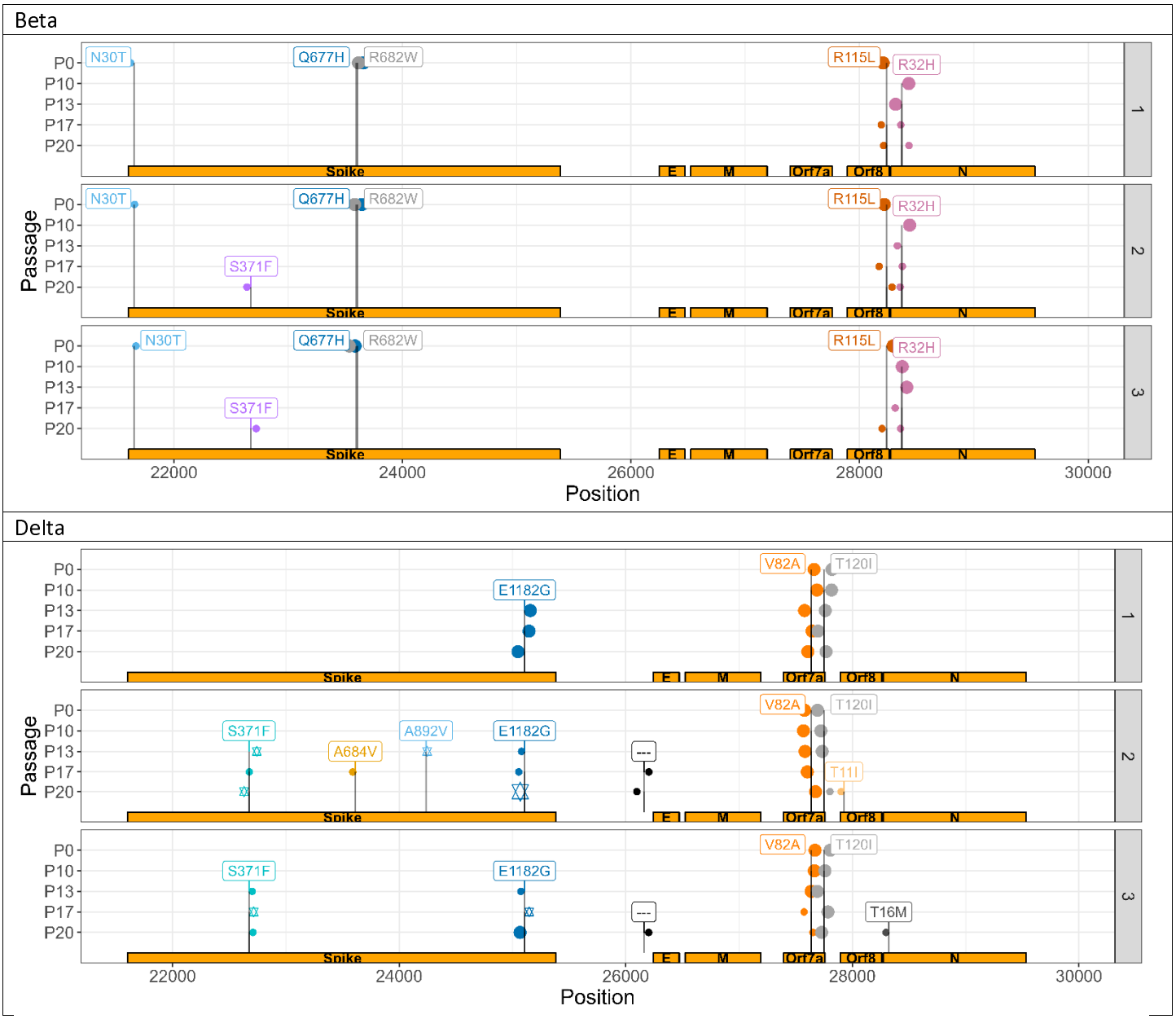


Figure 8. More variant alleles of clinical relevance arose de novo and persisted in spike for Delta lineage compared to Beta, with key variants being co-inherited. Each line and sample number refers to a select viral sample being passaged. P0 and P10 are identical across samples, given that it was passaged in a single animal up to P10. VAF: variant allele frequency. P refers to passage. Variant allele incidence by sample in Delta lineage virus. Larger points represent variant alleles with a within-sample, Illumina-derived allele-frequency of 1.0, smaller

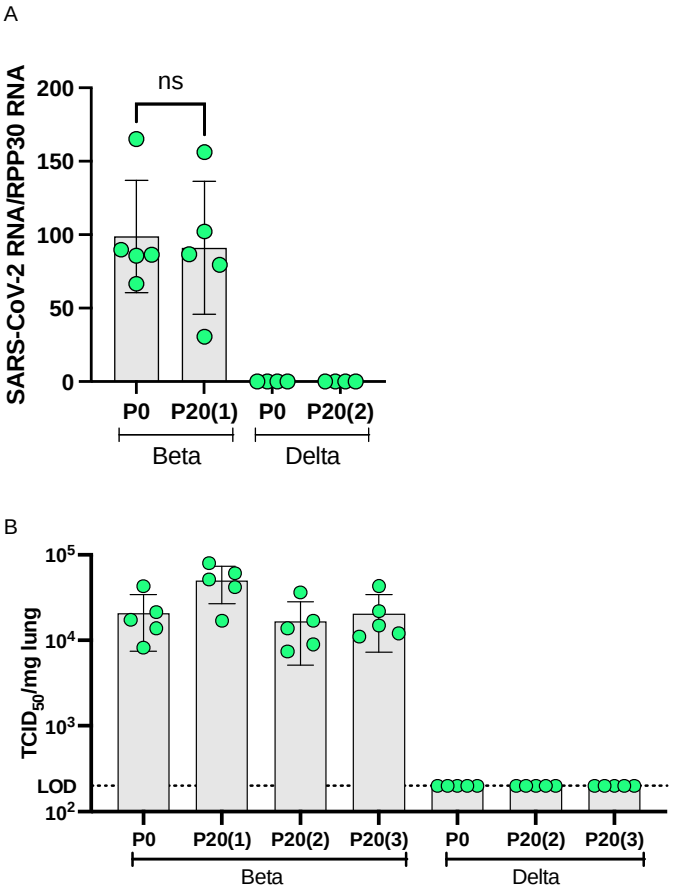
ones 0.5. Stars in a given passage represent alleles that are likely co-inherited per Nanopore sequencing.

485

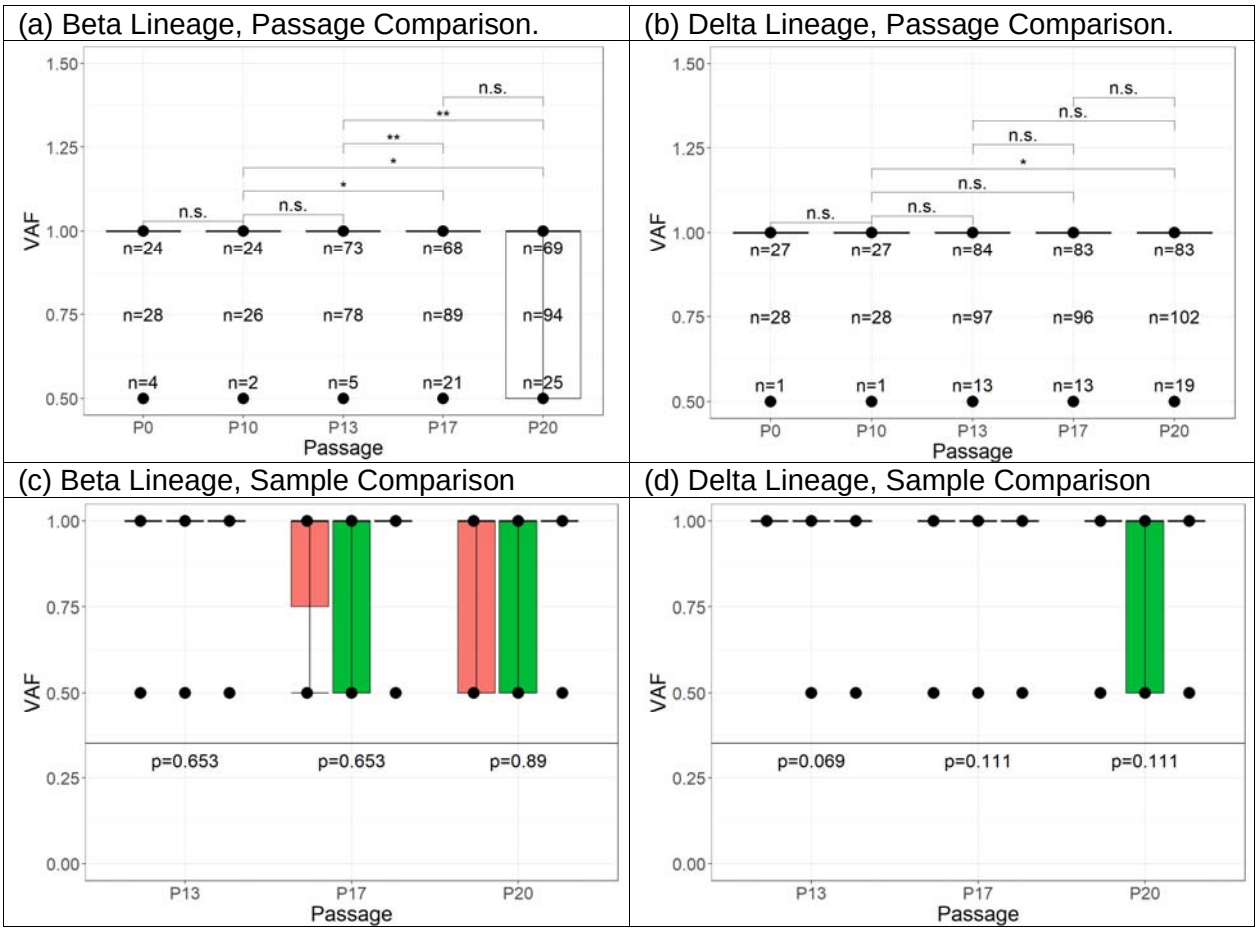
Table 1. Annotations of select variants that changed in frequency across the study.

Variant	Amino Acid Change	Annotation	Interacting Human Protein	Human protein implicated with COVID-19 severity or susceptibility
A929G	ORF1ab I222V, nsp2	Protein interacting*	GIGYF2, **FKBP15***, WASHC4, RAP1GDS1, POR, **eIF4E2***, SLC27A2	GIGYF2 (PMID 35878012)
C22674T	Spike S371F	Protein interacting, antibody escape, vaccine target, Omicron-associated variant allele[26]	GOLGA7, ZDHHC5	GOLGA7 (PMID 34961524), ZDHHC5 (PMID 34961524)
A25107G	Spike E1182G	Protein interacting, vaccine target	GOLGA7, ZDHHC5	GOLGA7 (PMID 34961524), ZDHHC5 (PMID 34961524)
G28237T	ORF8 R115L	Protein interacting*	COL6A1, PCSK6, **LOX***, **DNMT1***, NPC2, CISD3, ITGB1, PLAT, STC2, TOR1A, **PLOD2***, INHBE, CHPF2, UGGT2, FBXL12, PLEKHF2, SMOC1, POFUT1, **FKBP10***, ERLEC1, IL17RA, ADAMTS1, HS6ST2, SDF2, NEU1, GDF15, TM2D3, SIL1, EDEM3, ERP44, PVR, NGLY1, OS9, ADAM9, NPTX1, POGLUT2, POGLUT3, ERO1B, PLD3, FOXRED2, CHPF, PUSL1, HYOU1, MFGE8, FKBP7, GGH, EMC1	COL6A1 (PMID 35468151), **LOX** (PMID 34616409), PLAT (PMID 34786557), **PLOD2** (PMID 33328453), FBXL12 (PMID 34183789), **FKBP10** (PMID 35571107), IL17RA (PMID 33723527), HS6ST2 (PMID 32970989), NEU1 (PMID 36714013), GDF15 (PMID 35251002), ADAM9 (PMID 34698500), PLD3 (PMID 36182629)

Supplemental Figures



Supplementary Figure 1. Infection of C57BL6 mice with P0 and P20 Beta and Delta viruses. Mice (n = 5/group) were infected intranasally with 3000 $TCID_{50}$ of P0 and P20 Beta and Delta viruses. On the third day post infection, mice were euthanized and lungs collected for RNA extraction and infectious viral load determination. A) Mean +/- SD of normalized SARS-CoV-2 RNA copies as determined suing ddPCR. B) Mean $TCID_{50}$/mg of lung. LOD=limit of detection.



497 **Supplemental Figure 2. Boxplots of variant allele frequency by (a-b) passage and (c-d)**

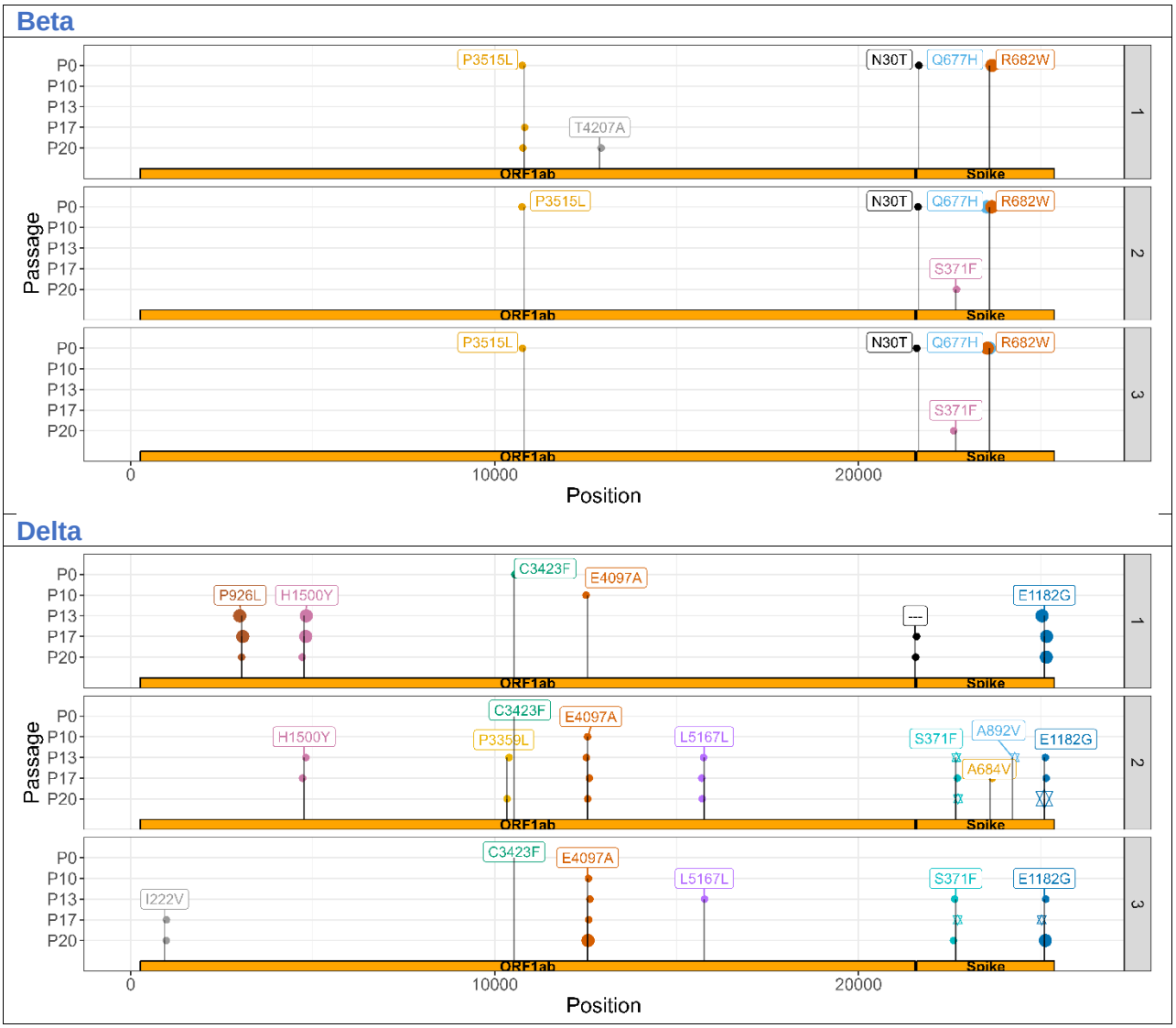
498 **sample.** Each N refers to the number of variant alleles for each MAF and in total against the

499 NC_045512.2 reference sequence. Single samples were sequenced for P0 and P10 with three

500 samples from P13, P17, and P20. Box fill in (c-d) signifies sample number. For significance,

501 n.s.: adjusted p-value > 0.05, *: adjusted p-value > 0.01, **: adjusted p-value ≤ 0.01, using

502 unpaired t-tests.



Supplemental Figure 3. More novel, high confidence, non-synonymous, changing alleles appeared in Delta versus Beta in the ORF1ab coding region. VAF: variant allele frequency. P refers to passage. Variant allele incidence by sample in Delta lineage virus. Larger points represent variant alleles with a within-sample, Illumina-derived allele-frequency of 1.0, smaller ones 0.5. Stars in a given passage represent alleles that are likely co-inherited per Nanopore sequencing.

Supplemental Table 1. Table of variant alleles that changed MAF during the study, alongside their annotations. Available in "Supplemental Tables.csv".

References:

1. Jacobs, J.L., Haidar, G. & Mellors, J.W. COVID-19: Challenges of Viral Variants. *Annu Rev Med* (2022).
2. V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **19**, 155-170 (2021).
3. Sun, F. *et al.* SARS-CoV-2 Quasispecies Provides an Advantage Mutation Pool for the Epidemic Variants. *Microbiol Spectr* **9**, e0026121 (2021).
4. Velavan, T.P. *et al.* Host genetic factors determining COVID-19 susceptibility and severity. *EBioMedicine* **72**, 103629 (2021).
5. Belsky, J.A. *et al.* COVID-19 in immunocompromised patients: A systematic review of cancer, hematopoietic cell and solid organ transplant patients. *J Infect* **82**, 329-338 (2021).
6. Van Cleemput, J. *et al.* Organ-specific genome diversity of replication-competent SARS-CoV-2. *Nat Commun* **12**, 6612 (2021).
7. Wang, R., Chen, J. & Wei, G.W. Mechanisms of SARS-CoV-2 Evolution Revealing Vaccine-Resistant Mutations in Europe and America. *J Phys Chem Lett* **12**, 11850-11857 (2021).
8. Liu, Y. *et al.* The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* **602**, 294-299 (2022).
9. Motozono, C. *et al.* SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe* **29**, 1124-1136 e11 (2021).
10. Cherian, S. *et al.* SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms* **9**(2021).
11. Liu, Y. *et al.* Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *Cell Rep* **39**, 110829 (2022).
12. Planas, D. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276-280 (2021).
13. Gu, H. *et al.* Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **369**, 1603-1607 (2020).
14. Wasik, B.R. *et al.* Influenza Viruses in Mice: Deep Sequencing Analysis of Serial Passage and Effects of Sialic Acid Structural Variation. *J Virol* **93**(2019).
15. Sun, S. *et al.* Characterization and structural basis of a lethal mouse-adapted SARS-CoV-2. *Nat Commun* **12**, 5654 (2021).
16. Huang, K. *et al.* Q493K and Q498H substitutions in Spike promote adaptation of SARS-CoV-2 in mice. *EBioMedicine* **67**, 103381 (2021).
17. Chung, H., Noh, J.Y., Koo, B.S., Hong, J.J. & Kim, H.K. SARS-CoV-2 mutations acquired during serial passage in human cell lines are consistent with several of those found in recent natural SARS-CoV-2 variants. *Comput Struct Biotechnol J* **20**, 1925-1934 (2022).
18. Sonnleitner, S.T. *et al.* The mutational dynamics of the SARS-CoV-2 virus in serial passages in vitro. *Virol Sin* **37**, 198-207 (2022).
19. Khalil, B.A., Elemam, N.M. & Maghazachi, A.A. Chemokines and chemokine receptors during COVID-19 infection. *Comput Struct Biotechnol J* **19**, 976-988 (2021).
20. Gadotti, A.C. *et al.* IFN-gamma is an independent risk factor associated with mortality in patients with moderate and severe COVID-19 infection. *Virus Res* **289**, 198171 (2020).
21. Berri, F. *et al.* Early plasma interferon-beta levels as a predictive marker of COVID-19 severe clinical events in adult patients. *J Med Virol* **95**, e28361 (2023).
22. Tada, T. *et al.* Convalescent-Phase Sera and Vaccine-Elicited Antibodies Largely Maintain Neutralizing Titer against Global SARS-CoV-2 Variant Spikes. *mBio* **12**, e0069621 (2021).

23. Dupont, L. *et al.* Neutralizing antibody activity in convalescent sera from infection in humans with SARS-CoV-2 and variants of concern. *Nat Microbiol* **6**, 1433-1442 (2021).
24. Tonkin-Hill, G. *et al.* Patterns of within-host genetic diversity in SARS-CoV-2. *Elife* **10**(2021).
25. Turakhia, Y. *et al.* Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. *Nature* **609**, 994-997 (2022).
26. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
27. Zeng, X.T., Yu, X.X. & Cheng, W. The interactions of ZDHHC5/GOLGA7 with SARS-CoV-2 spike (S) protein and their effects on S protein's subcellular localization, palmitoylation and pseudovirus entry. *Virol J* **18**, 257 (2021).
28. Wu, Z. *et al.* Palmitoylation of SARS-CoV-2 S protein is essential for viral infectivity. *Signal Transduct Target Ther* **6**, 231 (2021).
29. Zhang, Y., Qin, Z., Sun, W., Chu, F. & Zhou, F. Function of Protein S-Palmitoylation in Immunity and Immune-Related Diseases. *Front Immunol* **12**, 661202 (2021).
30. Cao, Y. *et al.* BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* **608**, 593-602 (2022).
31. Xu, Z. *et al.* SARS-CoV-2 impairs interferon production via NSP2-induced repression of mRNA translation. *Proc Natl Acad Sci U S A* **119**, e2204539119 (2022).
32. Song, L. *et al.* The main protease of SARS-CoV-2 cleaves histone deacetylases and DCP1A, attenuating the immune defense of the interferon-stimulated genes. *J Biol Chem* **299**, 102990 (2023).
33. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* **54**, 1 30 1-1 30 33 (2016).
34. Combadiere, B. *et al.* LOX-1-Expressing Immature Neutrophils Identify Critically-Ill COVID-19 Patients at Risk of Thrombotic Complications. *Front Immunol* **12**, 752612 (2021).
35. Pietzner, M. *et al.* Genetic architecture of host proteins involved in SARS-CoV-2 infection. *Nat Commun* **11**, 6397 (2020).
36. Wang, S. *et al.* Network Pharmacology and Bioinformatics Analyses Identify Intersection Genes of Vitamin D3 and COVID-19 as Potential Therapeutic Targets. *Front Pharmacol* **13**, 874637 (2022).
37. Staab-Weijnitz, C.A. *et al.* FK506-Binding Protein 10, a Potential Novel Drug Target for Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* **192**, 455-67 (2015).
38. Johnson, B.A. *et al.* Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293-299 (2021).
39. Rahimi, F. & Talebi Bezmin Abadi, A. Implications of the Emergence of a New Variant of SARS-CoV-2, VUI-202012/01. *Arch Med Res* **52**, 569-571 (2021).
40. Tarres-Freixas, F. *et al.* Heterogeneous Infectivity and Pathogenesis of SARS-CoV-2 Variants Beta, Delta and Omicron in Transgenic K18-hACE2 and Wildtype Mice. *Front Microbiol* **13**, 840757 (2022).
41. Zeng, C. *et al.* Neutralization of SARS-CoV-2 Variants of Concern Harboring Q677H. *mBio* **12**, e0251021 (2021).
42. Sui, Y., Li, J., Venzon, D.J. & Berzofsky, J.A. SARS-CoV-2 Spike Protein Suppresses ACE2 and Type I Interferon Expression in Primary Cells From Macaque Lung Bronchoalveolar Lavage. *Front Immunol* **12**, 658428 (2021).
43. Kim, K. *et al.* The roles of APOBEC-mediated RNA editing in SARS-CoV-2 mutations, replication and fitness. *Sci Rep* **12**, 14972 (2022).
44. Lei, C., Yang, J., Hu, J. & Sun, X. On the Calculation of TCID(50) for Quantitation of Virus Infectivity. *Virol Sin* **36**, 141-144 (2021).

45. Dubuc, I. *et al.* Cytokines and Lipid Mediators of Inflammation in Lungs of SARS-CoV-2 Infected Mice. *Front Immunol* **13**, 893792 (2022).
46. Lacasse, E. *et al.* SARS-CoV-2 Nsp2 Contributes to Inflammation by Activating NF-kappaB. *Viruses* **15**(2023).
47. Chen, S.-H., J Reiling, S., Quick, J. & Ragoussis, I. Artic nCoV-2019 McGill modified Lunascript Reverse Transcriptase sequencing protocol. *Protocols.io* (2020).
48. Freed, N.E., Vlkova, M., Faisal, M.B. & Silander, O.K. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biol Methods Protoc* **5**, bpaa014 (2020).
49. Freed, N. & Silander, O. SARS-CoV2 genome sequencing protocol (1200bp amplicon "midnight" primer set, using Nanopore Rapid kit) V.6. *Protocols.io* (2021).
50. J Reiling, S., Roy, A.-M., Chen, S.-H. & Ragoussis, I. nCoV-2019 McGill Nanopore LibPrep Protocol, 10 ng NB. *Protocols.io* (2020).
51. Garrison, E.P.M.G. Haplotype-based variant detection from short-read sequencing. *arXiv* (2012).
52. N'Guessan A, T.A., Sanchez-Quete F, Goitom E, Reiling SJ, Galvez JH, Nguyen TL, Nguyen HTL, Visentin F, Hachad M, Krylova K, Matthews S, Kraemer SA, Stretenowich P, Bourgey M, Djambazian H, Chen S, Roy A, Brookes B, Lee S, Simon M, Maere T, Vanrolleghem PA, Labelle M, Moreira S, Levade I, Bourque G, Ragoussis J, Dorner S, Frigon D, Shapiro BJ. Detection of prevalent SARS-CoV-2 variant lineages in wastewater and clinical sequences from cities in Québec, Canada. *medRxiv* (2022).
53. Gerstung, M. *et al.* Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nat Commun* **3**, 811 (2012).
54. Rao, R.S.P. *et al.* Evolutionary Dynamics of Indels in SARS-CoV-2 Spike Glycoprotein. *Evol Bioinform Online* **17**, 11769343211064616 (2021).
55. Ramachandran, V.K. *et al.* Comparison of variant callers for wastewater-based epidemiology. *medRxiv* (2022).
56. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* (2013).
57. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).