Long-term serial passaging of SARS-CoV-2 reveals signatures of convergent evolution

Charles S.P. Foster[1,2,*], Gregory J. Walker[1,2], Tyra Jean[1,2], Maureen Wong[1], Levent Brassil[3], Sonia Isaacs[1] , Yonghui Lyu[1] , Stuart Turville[4,5], Anthony Kelleher[4,5,6,] and William D. Rawlinson[1,2,3]

[1]Serology and Virology Division (SAViD), NSW Health Pathology, Prince of Wales Hospital, Sydney, NSW 2031, Australia

[2]School of Biomedical Sciences, Faculty of Medicine & Health, University of New South Wales, Sydney, NSW 2052, Australia

[3]School of Biotechnology and Biomolecular Sciences, Faculty of Science, University of New South Wales, Sydney, NSW 2052, Australia

[4]Kirby Institute, University of New South Wales, Sydney, NSW, Australia

[5]RNA Institute, University of New South Wales, Sydney, NSW, Australia

[6]St Vincent's Hospital, Sydney, New South Wales, Australia

[*]Correspondence: charles.foster@unsw.edu.au

Abstract

Measures to control the COVID-19 pandemic such as antiviral therapy and vaccination have been challenged by ongoing virus evolution under antiviral and immune pressures. Understanding viral evolutionary dynamics is crucial for responding to SARS-CoV-2, and preparing for the next pandemic, by informing prediction of virus adaptation, public health strategies, and design of broadly effective therapies. Whole-genome sequencing (WGS) of SARS-CoV-2 during the pandemic enabled fine-grained studies of virus evolution in the human population. Serial passaging *in vitro* offers a controlled environment to investigate the emergence and persistence of genetic variants that may confer selective advantage.

Nine virus lineages, including four “variants of concern” and three former “variants under investigation” as designated by the World Health Organisation, were chosen to investigate intra- and inter-lineage evolution through long-term serial passaging in Vero E6 cells. Viruses were sampled over at least 33 passages (range 33-100) and analysed using WGS to examine evolutionary dynamics and identify key mutations with implications for virus fitness, transmissibility, and immune evasion. All passages continued to replicate in culture, despite regular accumulation of mutations. There was evidence of convergent acquisition of mutations both across passage lines and compared with contemporaneous SARS CoV-2 clinical sequences from population studies. Some of these convergent mutations are hypothesised to be important in proliferation of SARS-CoV-2 lineages, such as by evading host immune responses (e.g. S:A67V, S:H655Y). Given these mutations arose *in vitro,* in the absence of a multicellular host immune response, this suggests virus genome mutation resulted from stochastic events, rather than immune-driven mutation. There was a regular gain and loss of low-frequency variants during serial passaging, but some became fixed in subsequent multiple passages, suggesting either a benefit of the mutation *in vitro*, or at least a lack of deleterious effect. Our findings reveal valuable insights into the evolution of SARS-CoV-2 by quantitatively investigating evolutionary dynamics of the virus over the greatest number of serial passages to date. Knowledge of these evolutionary trends will be useful for public health and the development of antiviral and vaccine measures to reduce the effects of SARS CoV-2 infection on the human population.

Introduction

The COVID-19 pandemic saw whole-genome sequencing (WGS) embraced on an unprecedented scale, with nearly 100 countries worldwide possessing SARS-CoV-2 WGS capability and contributing to publicly available sequencing repositories (Chen et al., 2022). Consequently, the number of high quality full-genome sequences for SARS-CoV-2 (>14,000,000 as of July 2023) far exceeds that of any other pathogen. The availability of these sequencing data have allowed the population-scale evolution of SARS-CoV-2 to be characterised with fine-grained detail, facilitating real-time monitoring of the spread of SARS-CoV-2, identification of new viral lineages, and have provided crucial insights into virus evolution. For example, the rise and fall of different SARS-CoV-2 lineages has been charted, with attempts to link the presence of key mutations with lineage success in promulgation through the human population. This wealth of genomic data has been vital in guiding public health responses, vaccine design, and antiviral treatments (Markov et al., 2023).

Global sequencing efforts have revealed that SARS-CoV-2 has continued to accumulate mutations throughout its genome since first emerging globally in 2020. The rate of mutation accumulation within SARS-CoV-2 (~1×10^{-6} to 2×10^{-6} mutations per nucleotide per replication cycle) is typical of betacoronaviruses, but below the rate typical in other RNA viruses that lack proofreading mechanisms (Markov et al., 2023). Charting the changes accruing in the SARS-CoV-2 genome in real time has allowed quantitative inference of evolutionary processes, such as the diversification of the virus into distinct phenotypes with different transmissibility, disease severity, and immune evasion properties (Markov et al., 2023). Retrospective consideration can divide the COVID-19 pandemic into several 'eras': the initial phase of the pandemic characterised by apparently limited evolution of the virus, the sudden emergence of the highly divergent virus lineages with altered phenotypes (variants of concern, VOCs) (Tay et al., 2022), and periods of gradual evolution within VOC lineages (Lythgoe et al., 2023; Tay et al., 2022). Determining the evolutionary forces responsible for each of these eras is key to understanding how the pandemic has progressed over time, how it will continue to progress as SARS-CoV-2 shifts towards being an endemic virus (Lavine et al., 2021), and how future pandemics might unfold.

Analysis of the changes that have occurred in SARS-CoV-2 genomes from the global clinical population suggests the predominant driver of evolution in SARS-CoV-2 has been the accumulation of mutations resulting in increased transmissibility (Markov et al., 2023). Increased transmissibility is a multifaceted trait resulting from a given virus having an increased ability to survive within an infected individual, shed from an infected individual, establish within a new individual, or some combination of each of these abilities. These are all influenced by virus evading the host immune response. A range of 'key mutations' that are partially and additively responsible for each of these components of transmissibility have been identified. For example, several mutations in the Spike region have been linked to enhanced receptor binding, such as D614G, identified early in the pandemic (Yurkovetskiy et al., 2020), and N501Y, which was identified in several VOCs (Harvey et al., 2021; Leung et al., 2021; Liu et al., 2022). These conclusions predominantly based on clinical observations, albeit with further laboratory-based exploration, are valuable for understanding how the pandemic has unfolded and will continue to unfold.

Whilst genomic data derived from routine genomic sequencing for clinical or surveillance purposes (henceforth referred to as the "clinical population" of SARS-CoV-2) has enhanced understanding the evolution of SARS-CoV-2, there are inherent limitations in relying solely on this information. The availability of WGS data is influenced by differential sequencing efforts between countries, especially when routine WGS is difficult to achieve in low- and middle-income countries (Chen et al., 2022), or when data sharing in online repositories is not consistent among jurisdictions. Additionally, the gain or loss of mutations in these populations is affected by multiple uncontrolled processes including natural selection, genetic drift, host immunity, and population

dynamics. Understanding the relative impact of each process is difficult in such an uncontrolled environment. Consequently, it is challenging to understand the adaptation of SARS-CoV-2 to specific selective pressures, or to predict the potential evolutionary pathways, when relying solely on the clinical population.

Serial passaging *in vitro* provides complementary insights into the ongoing evolution of viruses. In these experiments, the virus is passaged through successive generations within a cell line, allowing study of virus evolution in a controlled setting, as mutations accumulate over time during the course of passaging (Holland et al., 1982). Most commonly, *in vitro* serial passaging is conducted to develop virus stocks and assess therapeutics (Dimmock and Easton, 2014), or to assess virus attenuation for vaccine development (Minor, 2015). Serial passaging can also be used to study the evolutionary trajectory of viruses over long term experiments, with strictly controlled time periods (Andino and Domingo, 2015; Bertels et al., 2019). These experiments can determine the speed at which a virus develops resistance to antiviral therapies, or to understand how the virus evolves over longer periods in the absence of selective constraints including immune pressure.

Few studies have investigated the evolutionary dynamics of SARS-CoV-2 over time *in vitro*. Based on limited passaging of ancestral lineages (<15 passages), it was reported that SARS-CoV-2 accumulates mutations readily *in vitro*, especially when passaged in Vero E6 cells, with an estimated *in vitro* spontaneous mutation rate of $1.3 \times 10^{-6} \pm 0.2 \times 10^{-6}$ per-base per-infection cycle (Amicone et al., 2022). There is evidence for the convergent acquisition of mutations among samples taken through serial passage, such as through accumulation of mutations associated with Vero E6-cultured samples (Amicone et al., 2022), or associated with the global clinical population of SARS-CoV-2 (Chung et al., 2022). However, these studies explicitly focusing on tracking the evolution of SARS-CoV-2 *in vitro* generally studied a relatively small number of passages (e.g. <10), over a short time, and relatively few SARS-CoV-2 lineages. The greatest number of lineages published to date is 15 passages with two lineages from early in the pandemic (Amicone et al., 2022). Further *in vitro* serial passaging studies of SARS-CoV-2 are needed to: i) more closely approach the multiple passages occurring the human population, ii) assess sites of mutation fixation, iii) use this assessment to determine potential antiviral and vaccine resistance sites developing without antiviral or vaccine pressure, and iv) inform therapeutic and vaccine research.

In this study, we compared the accumulation of mutations *in vitro* in 11 different passaged viruses, corresponding to nine unique Pango lineages, throughout long-term serial passaging (33-100 passages per passaged virus line). The evolutionary dynamics of these passage lines was determined by characterising the mutations that accumulated, both in terms of their number and potential functional consequences. Additionally, we assessed mutations for evidence of convergence across passage lines and/or with key mutations from the global clinical population of SARS-CoV-2.

Methods

Rationale and study design

Since the beginning of the COVID-19 pandemic, routine whole-genome sequencing (WGS) has been conducted at the Virology Research Laboratory, Serology and Virology Division (SAViD), NSW Health Pathology East, on PCR-positive SARS-CoV-2 clinical specimens. Viruses from clinical specimens were isolated *in vitro*, and then serially passaged and sequenced to investigate the evolutionary dynamics of SARS-CoV-2 in culture (see additional methods below). Initially an early variant (A.2.2) was chosen for this purpose. Later, additional lineages were also grown *in vitro* based on their relevance within the clinical population globally, with a particular focus on 'variants of concern' (VOC) or 'variants under investigation' (VUI), as designated by the World Health Organisation. In total, 11 different passage lines were established for the purpose of this study, with the sample details outlined in Table 1. All samples were taken through a minimum of 33 passages, up to a maximum of 100 passages.

Table 1: Overview of the passage lines used within this study, including their nomenclatural assignments and the number of serial passages

Sample ID	GISAID Accession†	Pango Lineage	WHO Name*	Maximum Passage Number
POW003	EPI_ISL_707892	A.2.2	–	33
POW004	EPI_ISL_8767166	A.2.2	–	72
POW007	EPI_ISL_678338	A.2.2	–	66
POW005	EPI_ISL_18218201	B.1.319	–	100
nCoV_2318	EPI_ISL_1121976	P.1	Gamma	36
nCoV_2299	EPI_ISL_18218200	P.2	Zeta	36
UK002	EPI_ISL_794671	B.1.1.7	Alpha	57
RSA003	EPI_ISL_775245	B.1.351	Beta	57
RCoV0206	EPI_ISL_1660480	B.1.617.1	Kappa	39
nCoV_2605	EPI_ISL_1911187	B.1.617.2	Delta	39
RCoV1933	EPI_ISL_6814922	BA.1	Omicron (BA.1)	39

† All accessions here refer to consensus genomes from passage 0 (or the earliest available passage) that were uploaded to GISAID

* Names assigned by the World Health Organisation for Variants of Concern or Variants Under Investigation

Serial passaging

All SARS-CoV-2 cultures were performed in Vero E6s cells. Vero E6 cells were maintained in Minimum Essential Media supplemented with 10%-FBS and 1xPSG (MEM-10), and incubated at 37°C, 5% CO_2. The day prior to infection, a 24-well plate was seeded with 120,000 cells per well and incubated overnight, for approximately 90% confluency at the time of virus inoculation. The Vero E6 cell line was chosen as it supports viral replication to high titres due to its lack of interferon production and relatively high expression of the ACE-2 receptor (Emeny and Morgan, 1979). Therefore, using the Vero E6 cell line disentangles the effects of virus replication dynamics from host-mediated selective pressures. The appearance of any immune evasion-associated mutations

from the clinical population during in vitro passaging in Vero E6 cells can help to infer other fitness advantages of these mutations beyond immune evasion.

Clinical specimens (nasopharyngeal swabs in Virus Transport Media, referred to as "Passage 0") positive by diagnostic RT-qPCR were transported to a PC3 facility for virus culture. Specimens were spin-filtered (0.22 µm sterile Ultrafree-MC Centrifugal Filter) to remove cellular debris and reduce potential for bacterial and fungal contamination. Cell culture maintenance media was removed from wells, and 100 µL of virus-containing flow-through from spin-filtered specimens was added to wells. Plates were incubated for 1hr, before inoculum was removed and replaced with 500 µL maintenance media (MEM-2). Virus cultures were incubated for 3-4 days and observed for cytopathic effect (CPE). Cultured virus was then harvested (passage 1), diluted 1:1000, and used to inoculate a fresh plate of naïve Vero E6 cells (passage 2). Cultures of each SARS-CoV-2 lineage were maintained up to a maximum of passage 100. Aliquots of each virus passage were frozen at -80°C until later extraction for whole-genome sequencing.

Whole-genome sequencing

Samples were taken from each passage line and prepared for whole-genome sequencing (WGS). All wet lab methods follow those described previously (Foster et al., 2022), with the exception of the choice of amplicon scheme. In the present study, amplification was performed using the "Midnight" protocol, comprising tiled 1200 nucleotide amplicons (Freed et al., 2020), followed by 150 bp paired-end sequencing on an Illumina MiSeq. We chose not to sequence every passage of every passage line. Instead, we focused on sequencing passages 0–6, then every three passages onwards where possible.

Quality control and variant calling

Sequencing data was processed and analysed using an in-house bioinformatics pipeline ("CIS")(Foster, 2023), which leverages open-source bioinformatics tools for whole-genome sequencing, with workflow management controlled by Snakemake (Mölder et al., 2021). Description of the pipeline components relevant to this study are as follows. Briefly, initial QC of sequencing reads is performed with fastp v0.23.2 (Chen et al., 2018), with removal of residual sequencing adapters, error correction, and filtration of reads based on a quality score threshold. Reads passing QC are then mapped against the SARS-CoV-2 reference genome (NC_045512.2) using bwa mem v0.7.17-r1188 (Li, 2013), and amplicon primer regions are soft-clipped using ivar v1.3.1 (Grubaugh et al., 2019). Probabilistic realignment of reads and variant calling is performed with lofreq v2.1.5. A consensus genome is assembled using bcftools v1.13 (Danecek et al., 2021), and coverage statistics estimated using bedtools v2.30.0 (Quinlan and Hall, 2010). Finally, Pango lineages are assigned using pangolin (O'Toole et al., 2021) and nextclade (Aksamentov et al., 2021), then relevant pipeline statistics and additional QC metrics by nextclade are incorporated into a final report. The pipeline is freely available from https://github.com/charlesfoster/covid-illumina-snakemake.

Tracking mutational changes during serial passaging

We assessed the changes in called variants within and among passage lines throughout serial passaging using our in-house "vartracker" pipeline (see below, available from https://github.com/charlesfoster/vartracker), using the variant call format (VCF) files from lofreq as input. For each passage sample within a passage line, these VCF files were derived by calling variants against the SARS-CoV-2 reference genome (using the CIS pipeline, as per above), not against the prior passage number. Doing so allows variants to be discussed using standard notation for SARS-CoV-2, while also allowing the gain and/or loss of variants between passages to be tracked.

The vartracker pipeline comprises a series of in-house python scripts to allow interrogation of the mutations that accumulate within a passage line during serial passaging. During the pipeline, the VCF files for each passage line were prefiltered to retain variants with a minimum single-nucleotide variant (SNV) and indel frequency 0.03 and 0.1, respectively, using cyvcf2 (Pedersen and Quinlan, 2017) and bcftools. Functional annotation of called variants was inferred using bcftools csq (Danecek and McCarthy, 2017), followed by merging into a single file. The merged VCF was then processed using custom python code to track the change in presence/absence of a given variant throughout the trajectory of the passage line (from passage 0 to the final passage), as well as the change in allele frequency of each variant throughout passaging. Variants were classified under four main broad categories: present in passage 0 but lost throughout passaging ('original_lost'), present in passage 0 and retained throughout passaging ('original_retained'), absent in passage 0 but gained and retained throughout passaging ('new_persistent'), or absent in passage 0 but gained and lost throughout passaging ('new_transient'). All variants were also subject to a suite of quality control measures incorporating the estimated variant depth, the overall site depth, the depth in a window of sites surrounding the variant site, and an overall metric.

We investigated convergent evolution of any new_persistent or new_transient variants (a) among passage lines, and (b) compared to the clinical population of SARS-CoV-2 globally by searching them against a database of SARS-CoV-2 mutations comprising information on the functional consequences of mutations with associated references to primary literature (Anwar et al., 2022). Additionally, we calculated the rate of mutation accumulation within each passage line, considering the rate among the overall genome as well as individual genes/regions within the genome, using the following formula: $M_a(r) = \frac{\Sigma f_r}{P.L_r}$(following the approach of Amicone et al., 2022). In this formula, Σf_r is the sum of the frequency of all detected mutations in a region r, P is the number of passages, and L_r is the length of r. We calculated the rate for each applicable passage up to and including passage 33 to observe whether the inferred mutation rate varies over time.

We investigated the evolutionary dynamics within all passage lines from passage 0, or the earliest passage successfully sequenced, until their final passage number (Table 1). However, for all comparisons among passage lines, we restricted our focus to passages 0–33, with passage 33 representing the maximum passage number in common that all passage lines had reached. We also investigated in further detail the evolutionary trajectory of passage line "POW005" (Pango lineage: B.1.319), which has been maintained for the greatest number of passages (100) across all passage lines. The full results for all other passage lines across all passages are not discussed in this manuscript, but are available within the supplementary materials (Table S2). It is also worth noting that sequencing data for some passage numbers from some passage lines are not available. In these cases, particularly with samples corresponding to VOCs, there was insufficient material from the early passages to conduct WGS and/or the material was needed for unrelated neutralization assay studies.

Phylogenetic analysis

A phylogenetic tree was estimated to investigate further the evolution of the passaged virus lines *in vitro*. All isolate consensus genomes were aligned against the Wuhan Hu-1 reference genome using minimap2 (Li, 2018), followed by trimming of the 5' and 3' UTRs and conversion to multifasta format using gofasta (Jackson, 2022). We then used an in-house script to mask known problematic sites from the alignment (De Maio et al., 2020), and filtered out sites with too many gaps using esl-alimask (hmmer.org). Finally, we inferred a maximum likelihood phylogeny for the samples using iqtree2 (Nguyen et al., 2015), with the best fitting nucleotide substitution model chosen using a constrained search using ModelFinder ('mset': GTR, TIM2, GTR; 'mfreq': F; 'mrate': 'G,R')

(Kalyaanamoorthy et al., 2017), and support estimated using 1000 ultrafast bootstraps (Hoang et al., 2018).

Results

Sequencing metrics

The majority of samples from each passage line were successfully sequenced (Supplementary Table S1). Excluding one sample that yielded no sequence, the mean reference genome coverage at a depth of at least 10 reads was 99.19% (SD: 1.07). Each of these samples was also sequenced at a high mean depth (mean 1334.48, SD: 753). All except one passage line were called as the same lineage as their original isolate (passage 0) after their maximum number of passages. The exception was “POW005”, which was originally called as the lineage B.1.319, but from passage 5 onwards was called as B.1 by pangolin, albeit with conflicting placements by Usher (B.1(1/3); B.1.319(1/3); B.1.616(1/3)). However, nextclade called the lineage of passage POW005 as B.1.319 from passage 0 through to passage 100, despite the accumulation of many private mutations.

Evolutionary dynamics across passage lines *in vitro*

All passage lines accumulated a range of mutations throughout the breadth of the genome (Figure 1). Many of these mutations were detected at a very low variant allele frequency (VAF) within a given passage (e.g., <5%), and were then lost in subsequent passages. Conversely, some new mutations that arose during passaging persisted until the final passage, and became near-fixed at a consensus-level frequency (Supplementary Table S2. Supplementary Figure S1). Most variants (~90%) relative to the Wuhan Hu-1 reference genome present in passage 0 of a given passage line were retained during the course of serial passaging. Occasionally these ‘original’ variants were lost, most commonly when they were originally present at a sub-consensus level, but sometimes even when well supported (allele frequency >50%, ~2% of all variants). Variants were detected in all genes of the SARS-CoV-2 genome, but occurred at different frequencies among genes and among passage lines (Figure 2, Supplementary Figure S1). The inferred mutation rates *in vitro* varied among passage lines, and varied within passage lines over time (Supplementary Figure S2).

The progressive accumulation of mutations throughout the course of serial passaging was evident in the inferred phylogenetic tree (Supplementary Figure S3). All samples within a given passage line form clades exclusive of the other passage lines, demonstrating that the convergent acquisition of some mutations among passage lines (as discussed below) did not impact upon inferred phylogenetic relationships. Overall, the phylogenetic relationships among passage lines reflects the expected topology based on the relationships among SARS-CoV-2 lineages within the global phylogeny.

Mutation accumulation over 100 serial passages

In the “POW005” passage line (Pango lineage: B.1.319) there was a continuous accumulation of mutations throughout the course of serial passaging, with no signs of an asymptotic “levelling off” of mutation accumulation (Figure 3, Supplementary Table S2). The initial (passage 0) isolate possessed 10 variants relative to the Wuhan Hu-1 reference genome: one in the 5’ UTR and nine occurring in coding regions, of which eight were present as the majority allele and seven were non-synonymous variants. Throughout the course of serial passaging, POW005 accumulated 131 new variants (58 missense including two in-frame deletions, 63 synonymous, 10 in intergenic regions/UTRs), of which 97 were transient and 34 persisted until passage 100. The sole low-frequency variant detected in passage 0 was lost during passaging, although the other 10 variants present at passage 0 were retained.

Signatures of convergent evolution

Although the majority of variants arising during serial passaging were passage-line specific, there were also many cases where different passage lines independently acquired the same variants in a convergent manner. Most instances of convergence were restricted to variants arising in only a small number of passage lines, and/or were occurring only at a low allele frequency. However, other variants were acquired and went to near-fixation in several to many lineages (Supplementary Table S2, Supplementary Figure S1). At least 18 different variants that arose during serial passaging have been previously reported in the scientific literature as having important functional consequences in the global clinical population of SARS-CoV-2 (Table 2, Supplementary Table S3). The classification of these variants falls under 20 unique categories within a SARS-CoV-2 functional mutation literature compendium (Anwar et al., 2022).

Discussion

After more than three and a half years of the COVID-19 pandemic, SARS-CoV-2 continues to diversify at a remarkable rate in the global clinical population. We found that SARS-CoV-2 steadily accumulated mutations *in vitro* throughout most of its genome, reflecting what continues to occur in the clinical population (Figure 1, Supplementary Table S2). The accumulation of mutations over time was reflected in a gradual divergence between passage 0 and the final passage *within* each passage line, but without significantly affecting the phylogenetic relationships between passage lines (Figure S3). The greatest mutation rate occurred in the four main VOC lineages (Delta, Omicron BA.1, Alpha, Beta) (Figure S2). There was a more than threefold difference in mutation rate between the passage lines with the greatest (Delta; ~8.51 x 10^{-05} mutations per passage per nucleotide) and smallest ("POW004"; ~2.64 x 10^{-05} mutations per passage per nucleotide) mutation rates.

Previous studies have concluded that the spike region of SARS-CoV-2 is a mutational hotspot. Whilst we observed many mutations arising in the spike gene, after correcting for gene length, the number of mutations in spike was comparable with that of other genes (Figure 2). Genes coding for other structural and accessory proteins (e.g. M, ORF7a, ORF7b) accumulated mutations at a greater rate than other regions, including spike, with variation between passage lines (Figure S2). While the spike region is clearly an important region of the SARS-CoV-2 genome for mutation, our results further reflect recent studies highlighting the importance of other regions of the genome for understanding the ongoing evolution of SARS-CoV-2 (Arshad et al., 2023; Foster et al., 2023; Redondo et al., 2021; Zandi et al., 2022), and the capacity of the virus to adapt to new evolutionary pressures such as new drug targets.

The success of the various lineages of SARS-CoV-2 in promulgating through the human population has been linked to functional effects of mutations that accrue over time (Markov et al., 2023). Determining why some mutations go on to be fixed, whereas others are temporary, is important to understand the evolutionary trajectory of SARS-CoV-2. In general, positive selection has played an important role in the COVID-19 pandemic, for example with mutations aiding in immune escape being selected for in a chronically infected host, particularly if that host is immune compromised. This process is hypothesised to have led to the emergence of variants of concern such as Omicron (Viana et al., 2022). Positive selection of mutations has been observed to arise during serial passaging in cell lines (Amicone et al., 2022), as well as in other animal models such as minks or domestic cats (Bashor et al., 2022; Shuai et al., 2020). We observed many variants newly arising in all passaged SARS-CoV-2 lineages (Supplementary Table S2, Supplementary Figure S1). Some of the >170 newly arising variants were transient and purged from the virus populations throughout the course of serial passaging, but some persisted until the final serial passages, occasionally at (near-)fixation within the virus populations. One signature of positive selection is a selective sweep, whereby a rare or previously non-existing allele rapidly increases in frequency within a population. Therefore, we considered the fixation of these newly arising mutations throughout the course of serial passaging to represent possible positive selection for mutations beneficial to virus survival *in vitro.*

Another hallmark of the COVID-19 pandemic has been high levels of homoplasy most likely caused by convergent and parallel evolution. Even from early within the pandemic, sites within the SARS-CoV-2 genome were noted as being highly homoplastic and recommended to be masked from alignments when estimating phylogenies (De Maio et al., 2020). As the pandemic has progressed, multiple mutations speculated to impart fitness benefits on SARS-CoV-2 have been observed. For example, the spike mutations N501Y, E484K and ΔH69/V70 have arisen several times independently in lineages designated as variants of concern, and are proposed to lead to increased immune escape and/or increased ACE2 binding affinity (Harvey et al., 2021). Therefore,

to understand further the impact of the mutations arising *de novo* during long-term passaging, we determined whether there were any signatures of convergent evolution. We were especially interested in whether mutations that are clinically significant, for example because of their noted impacts on transmissibility or disease severity, would arise throughout the course of serial passaging.

Many of the same *de novo* mutations arose convergently across passage lines throughout the course of serial passaging (Supplementary Table S3). Generally these mutations only appeared at a low frequency in few passage lines, but occasionally went to fixation in multiple passage lines, most notably with S:H655Y (see below). Some convergent mutations were expected to arise given that they have been noted repeatedly to arise during serial passaging in Vero E6 cells, especially in the region of the Spike furin-cleavage site (Ogando et al., 2020; Sasaki et al., 2021). For example, the S:YQTQTN674Y mutation, a deletion of five amino acids near the spike S1/S2 cleavage site, arose independently in POW003 (A.2.2), POW005 (B.1.319) and POW007 (A.2.2). This same mutation emerged multiple times in two background lineages in a previous study (Amicone et al., 2022).

We also considered instances of convergent evolution with the broader clinical population of SARS-CoV-2. Comparing our results with a curated database of SARS-CoV-2 mutations revealed that there were many such cases of convergent evolution (Table 2). Noted functional impacts of the newly arising mutations in our study are related to various aspects of pharmaceutical effectiveness, including vaccine efficacy and antibody neutralization (Table 2). Examples include the spike mutations L5F, A67V, T76I, A222V and others (Table 2). Likewise, the nsp14:A504V mutation that arose in our passage line of the Alpha variant has been demonstrated clinically to temporally associate with remdesivir resistance (Gandhi et al., 2022). The repeated evolution of these clinically significant mutations conclusively demonstrates that variants that are related to immune evasion can arise during *in vitro* serial passaging even in the absence of immune or drug -related selective pressures. This is because (a) we passaged the viruses through a Vero E6 cell line, which lacks an innate immune response, and (b) we did not directly apply any selective agents to the samples during passaging, such as neutralising antibodies. This result raises the possibility that (at least some) mutations associated with immune escape that arose in the clinical population were not driven by intra-host selection in the presence of (e.g.) antiviral therapies (cf. Weigang et al., 2021), but instead arose *de novo* with intrinsic immune escape benefits. However, it is possible that there could be another benefit of these putative immune-related mutations to the survival of SARS-CoV-2 outside of escaping an immune response.

Alternatively, the proliferation of these mutations *in vitro* could be explained by a relaxation of negative selection, which can lead to the chance success of alleles through genetic drift (Lahti et al., 2009), rather than by positive selection. Consequently, many of the putative immune-related mutations we observed during this study could have arisen and been successful through chance alone, such as through transmission bottlenecks between successive serial passages. Analogous situations in the global clinical populations of SARS-CoV-2 have been observed, whereby mutations that were either non-beneficial, or potentially even mildly deleterious, became fixed within given regions after public health measures created transmission bottlenecks (e.g. Foster et al., 2023). Likewise, the ORF8 accessory region remains a hotspot for mutations throughout the COVID-19 pandemic (Arduini et al., 2023; Hisner et al., 2023), possibly as a result of a relaxation of purifying selection, as was hypothesised for SARS-CoV (Zinzula, 2021).

The most striking example of convergent evolution of a clinically significant mutation across passage lines was the repeated evolution of S:H655Y. This mutation was already present in passage 0 in two passage lines, and arose newly throughout serial passaging in six of the remaining nine passage lines. The S:H655Y mutation was originally detected in early cases of the

Gamma and Omicron VOCs. Impacts of the S:H655Y mutation include acting as a “gateway” mutation by increasing the fitness of lineages with complex combinations of clinically significant mutations, and increased spike cleavage and fusogenicity associated with enhancement of the endosomal cell entry pathway (Yamamoto et al., 2022; Yurkovetskiy et al., 2020). A putative switch towards preferentially using an endosomal cell entry pathway within our passage lines, rather than the ACE2-TMPRSS2 pathway typical of most variants prior to omicron (Jackson et al., 2022), is not surprising given that Vero E6 cell lines do not express TMPRSS2. Similar changes in virus tropism have been observed in the clinical population. For example, several lineages of the Omicron BA.1 and BA.2 variants (and their descendants) used an endosomal cell entry pathway rather than the ACE2-TMPRSS2 pathway for virus entry (Aggarwal et al., 2022). This change in tropism resulted in Omicron replicating more efficiently in the bronchi, but not productively infecting human alveolar type 2 cells of the lung (Hui et al., 2022; Lamers et al., 2022). These changes, among other factors, likely contributed to the high transmissibility and decreased pathogenicity observed clinically since the emergence of Omicron and its sub-lineages (Hyams et al., 2023; Yuan et al., 2022).

In addition to virus tropism-associated changes, any mutations associated with enhanced virus replication *in vitro* would be expected to be beneficial for virus survival and selected for, as is the case with replication-enhancing mutations in the clinical population. For example, the emergence of the spike mutation D641G early in the COVID-19 pandemic was linked with the proliferation of virus with enhanced replication (Zhou et al., 2021). Additional changes in the spike region have been linked to enhanced virus replication at subsequent stages of the pandemic, such as the P681R mutation within the Delta VOC (Saito et al., 2022). However, the impact of mutations in other genomic regions on virus replication is also important to consider. A well characterised pair of sequential non-synonymous amino acid changes in the nucleocapsid gene, N:R203K and N:G204R, augment virus replication in SARS-CoV-2 (Johnson et al., 2022). One example of a replication-associated mutation that arose during our long-term passaging study is P203L within nsp14, which encodes a proofreading exonuclease. This mutation has been linked previously to altered replication (Takada et al., 2023).

Early in the pandemic, the furin cleavage site (PRRAR) within the S1/S2 domain of the SARS-CoV-2 spike protein was noted as an unusual characteristic of the virus compared to other coronaviruses in the same clade (Coutard et al., 2020). The furin cleavage site has been proposed as a key component of the pathogenesis of SARS-CoV-2, playing an important role in the affinity of for human hosts (Johnson et al., 2021). Deletion of the furin cleavage site has been linked with a fitness advantage in Vero E6 cells, albeit with this advantage reduced when ectopic expression of TMPRSS2 is introduced to the Vero E6 system (Johnson et al., 2021). Our results further strengthen these previous findings, with the furin cleavage site being a hotspot for *de novo* mutations on culture *in vitro* (Table S2). Four different missense amino acid replacements were observed at amino acid position 682 of spike (R682G, R682W, R682L, R682Q) at varying allele frequencies, as well as a six amino acid deletion in one passage line (RARSVA683R). All but one of these mutations appeared within four serial passages. SARS-CoV-2 clearly adapts readily and rapidly to its environment to enable continued virus proliferation.

We also observed the loss of at least one clinically significant mutation throughout the course of serial passaging. One key SNP hypothesised to have contributed to the success of the Omicron BA.1 lineage is S:S375F, a mutation within the receptor-binding domain of the Spike protein responsible for attenuation of spike cleavage efficiency and fusogenicity as well as decreased ACE2 binding affinity (Kimura et al., 2022). This key mutation was present in our BA.1 passage line at a VAF of ~1 from passage 0 until passage 6 (i.e., fixed as the majority allele), but was only present at a low VAF of 13% in passage 9, and was then not detected in subsequent

passages (Supplementary Table S2). The dominance of S:S375F in global Omicron sequences suggests the site may have been subject to initial positive selection followed by purifying selection to retain benefits imparted by the mutation. However, its loss throughout serial passaging suggests that the benefits in the clinical population do not translate to *in vitro* benefits, at least in Vero E6 cell lines.

Apart from revealing signatures of convergence with previously noted mutations of interest, our results provide a resource for investigating future changes occurring in SARS-CoV-2 globally. Any mutations arising convergently among passage lines that have not been noted previously as occurring commonly during serial passaging, or in the global clinical population, are worth further investigation. The repeated appearance of these mutations during *in vitro* culture implies they benefit SARS-CoV-2, at least *in vitro*, and could also add to the growing list of Vero E6-associated mutations in SARS-CoV-2. The same is true for other mutations arising *de novo* even in a single passage line.

Data availability

The consensus genomes of all passage 0 samples from this study have been uploaded to GISAID (see Table 1), and all sequence data have been uploaded to NCBI's GenBank Sequence Read Archive under accession number PRJNA1018257. The pipeline to process and analyse raw sequencing reads is available from https://github.com/charlesfoster/covid-illumina-snakemake, and the pipeline to summarise longitudinal variant calling results is available from https://github.com/charlesfoster/vartracker.

References

Aggarwal, A., Akerman, A., Milogiannakis, V., Silva, M.R., Walker, G., Kidinger, A., Angelovich, T., Waring, E., Amatayakul-Chantler, S., Roth, N., Coppola, G., Yeang, M., Jean, T., Foster, C., Hoppe, A.C., Munier, C.M.L., Darley, D., Churchill, M., Starck, D., Christ, D., Matthews, G., Rawlinson, W., Kelleher, A.D., Turville, S., 2022. SARS-CoV-2 Omicron BA.5: Evolving tropism and evasion of potent humoral responses and resistance to clinical immunotherapeutics relative to viral variants of concern. medRxiv 2022.07.07.22277128. https://doi.org/10.1101/2022.07.07.22277128

Aksamentov, I., Roemer, C., Hodcroft, E.B., Neher, R.A., 2021. Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software 6, 3773. https://doi.org/10.21105/joss.03773

Amicone, M., Borges, V., Alves, M.J., Isidro, J., Zé-Zé, L., Duarte, S., Vieira, L., Guiomar, R., Gomes, J.P., Gordo, I., 2022. Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. Evolution, Medicine, and Public Health 10, 142–155. https://doi.org/10.1093/emph/eoac010

Andino, R., Domingo, E., 2015. Viral quasispecies. Virology, 60th Anniversary Issue 479–480, 46–51. https://doi.org/10.1016/j.virol.2015.03.022

Anwar, M.Z., Gill, I.S., Iseminger, M., Sehar, A., Igwacho, K.D., Vora, K., Domselaar, G.V., Gordon, P.M.K., Hsiao, W.W., 2022. COVID-MVP: an interactive visualization for tracking SARS-CoV-2 mutations, variants, and prevalence, enabled by curated functional annotations and portable genomics workflow. bioRxiv 2022.06.07.493653. https://doi.org/10.1101/2022.06.07.493653

Arduini, A., Laprise, F., Liang, C., 2023. SARS-CoV-2 ORF8: A Rapidly Evolving Immune and Viral Modulator in COVID-19. Viruses 15, 871. https://doi.org/10.3390/v15040871

Arshad, N., Laurent-Rolle, M., Ahmed, W.S., Hsu, J.C.-C., Mitchell, S.M., Pawlak, J., Sengupta, D., Biswas, K.H., Cresswell, P., 2023. SARS-CoV-2 accessory proteins ORF7a and ORF3a use distinct mechanisms to down-regulate MHC-I surface expression. Proceedings of the National Academy of Sciences 120, e2208525120. https://doi.org/10.1073/pnas.2208525120

Bashor, L., Gagne, R.B., Bosco-Lauth, A., Stenglein, M., VandeWoude, S., 2022. Rapid evolution of SARS-CoV-2 in domestic cats. Virus Evolution 8, veac092. https://doi.org/10.1093/ve/veac092

Bertels, F., Leemann, C., Metzner, K.J., Regoes, R.R., 2019. Parallel Evolution of HIV-1 in a Long-Term Experiment. Molecular Biology and Evolution 36, 2400–2414. https://doi.org/10.1093/molbev/msz155

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, Z., Azman, A.S., Chen, X., Zou, J., Tian, Y., Sun, R., Xu, X., Wu, Y., Lu, W., Ge, S., Zhao, Z., Yang, J., Leung, D.T., Domman, D.B., Yu, H., 2022. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. Nat Genet 54, 499–507. https://doi.org/10.1038/s41588-022-01033-y

Chung, H., Noh, J.Y., Koo, B.-S., Hong, J.J., Kim, H.K., 2022. 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. Computational and Structural Biotechnology Journal 20, 1925–1934. https://doi.org/10.1016/j.csbj.2022.04.022

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G., Decroly, E., 2020. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage

site absent in CoV of the same clade. Antiviral Research 176, 104742. https://doi.org/10.1016/j.antiviral.2020.104742
Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10. https://doi.org/10.1093/gigascience/giab008
Danecek, P., McCarthy, S.A., 2017. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics 33, 2037–2039. https://doi.org/10.1093/bioinformatics/btx100
De Maio, N., Walker, C., Borges, R., Weilguny, L., Slodkowicz, G., Goldman, N., 2020. Issues with SARS-CoV-2 sequencing data [WWW Document]. Virological. URL https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473 (accessed 4.14.21).
Dimmock, N.J., Easton, A.J., 2014. Defective Interfering Influenza Virus RNAs: Time To Reevaluate Their Clinical Potential as Broad-Spectrum Antivirals? Journal of Virology 88, 5217–5227. https://doi.org/10.1128/jvi.03193-13
Emeny, J.M., Morgan, M.J., 1979. Regulation of the interferon system: evidence that Vero cells have a genetic defect in interferon production. J Gen Virol 43, 247–252. https://doi.org/10.1099/0022-1317-43-1-247
Foster, C.S.P., 2023. covid-illumina-snakemake: a pipeline to facilitate genomic surveillance of SARS-CoV-2. https://doi.org/10.5281/zenodo.8253928
Foster, C.S.P., Bull, R.A., Tedla, N., Santiago, F., Agapiou, D., Adhikari, A., Walker, G.J., Shrestha, L.B., Van Hal, S.J., Kim, K.W., Rawlinson, W.D., 2023. Persistence of a Frameshifting Deletion in SARS-CoV-2 ORF7a for the Duration of a Major Outbreak. Viruses 15, 522. https://doi.org/10.3390/v15020522
Foster, C.S.P., Stelzer-Braid, S., Deveson, I.W., Bull, R.A., Yeang, M., Au, J.-P., Ruiz Silva, M., van Hal, S.J., Rockett, R.J., Sintchenko, V., Kim, K.W., Rawlinson, W.D., 2022. Assessment of Inter-Laboratory Differences in SARS-CoV-2 Consensus Genome Assemblies between Public Health Laboratories in Australia. Viruses 14, 185. https://doi.org/10.3390/v14020185
Freed, N.E., Vlková, M., Faisal, M.B., Silander, O.K., 2020. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. Biology Methods and Protocols 5, bpaa014. https://doi.org/10.1093/biomethods/bpaa014
Gandhi, S., Klein, J., Robertson, A.J., Peña-Hernández, M.A., Lin, M.J., Roychoudhury, P., Lu, P., Fournier, J., Ferguson, D., Mohamed Bakhash, S.A.K., Catherine Muenker, M., Srivathsan, A., Wunder, E.A., Kerantzas, N., Wang, W., Lindenbach, B., Pyle, A., Wilen, C.B., Ogbuagu, O., Greninger, A.L., Iwasaki, A., Schulz, W.L., Ko, A.I., 2022. De novo emergence of a remdesivir resistance mutation during treatment of persistent SARS-CoV-2 infection in an immunocompromised patient: a case report. Nat Commun 13, 1547. https://doi.org/10.1038/s41467-022-29104-y
Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., Gurfield, N., Van Rompay, K.K.A., Isern, S., Michael, S.F., Coffey, L.L., Loman, N.J., Andersen, K.G., 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biology 20, 8. https://doi.org/10.1186/s13059-018-1618-7
Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S.J., Robertson, D.L., 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol 19, 409–424. https://doi.org/10.1038/s41579-021-00573-0
Hisner, R., Gueli, F., Peacock, T., 2023. Repeated loss of ORF8 expression in circulating SARS-CoV-2 lineages - SARS-CoV-2 coronavirus. Virological.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35, 518–522. https://doi.org/10.1093/molbev/msx281

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., VandePol, S., 1982. Rapid Evolution of RNA Genomes. Science 215, 1577–1585. https://doi.org/10.1126/science.7041255

Hui, K.P.Y., Ho, J.C.W., Cheung, M., Ng, K., Ching, R.H.H., Lai, K., Kam, T.T., Gu, H., Sit, K.-Y., Hsin, M.K.Y., Au, T.W.K., Poon, L.L.M., Peiris, M., Nicholls, J.M., Chan, M.C.W., 2022. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. Nature 603, 715–720. https://doi.org/10.1038/s41586-022-04479-6

Hyams, C., Challen, R., Marlow, R., Nguyen, J., Begier, E., Southern, J., King, J., Morley, A., Kinney, J., Clout, M., Oliver, J., Gray, S., Ellsbury, G., Maskell, N., Jodar, L., Gessner, B., McLaughlin, J., Danon, L., Finn, A., Morley, A., Langdon, A., Turner, A., Mattocks, A., Osborne, B., Grimes, C., Mitchell, C., Adegbite, D., Bridgeman, E., Scott, E., Perkins, F., Bayley, F., Ruffino, G., Valentine, G., Tilzey, G., Campling, J., Wright, J.K., Brzezinska, J., Cloake, J., Milutinovic, K., Helliker, K., Maughan, K., Fox, K., Minou, K., Ward, L., Fleming, L., Morrison, L., Smart, L., Wright, L., Grimwood, L., Bellavia, M., Clout, M., Vasquez, M., Gonzalez, M.G., Jeenes-Flanagan, M., Chang, N., Grace, N., Manning, N., Griffiths, O., Croxford, P., Sequenza, P., Lazarus, R., Walters, R., Marlow, R., Heath, R., Antico, R., Arachchge, S.N., Suppiah, S., Mona, T., Riaz, T., Mackay, V., Maseko, Z., Taylor, Z., Friedrich, Z., Szasz-Benczur, Z., 2023. Severity of Omicron (B.1.1.529) and Delta (B.1.617.2) SARS-CoV-2 infection among hospitalised adults: a prospective cohort study in Bristol, United Kingdom. The Lancet Regional Health – Europe 25. https://doi.org/10.1016/j.lanepe.2022.100556

Jackson, B., 2022. gofasta: command-line utilities for genomic epidemiology research. Bioinformatics 38, 4033–4035. https://doi.org/10.1093/bioinformatics/btac424

Jackson, C.B., Farzan, M., Chen, B., Choe, H., 2022. Mechanisms of SARS-CoV-2 entry into cells. Nat Rev Mol Cell Biol 23, 3–20. https://doi.org/10.1038/s41580-021-00418-x

Johnson, B.A., Xie, X., Bailey, A.L., Kalveram, B., Lokugamage, K.G., Muruato, A., Zou, J., Zhang, X., Juelich, T., Smith, J.K., Zhang, L., Bopp, N., Schindewolf, C., Vu, M., Vanderheiden, A., Winkler, E.S., Swetnam, D., Plante, J.A., Aguilar, P., Plante, K.S., Popov, V., Lee, B., Weaver, S.C., Suthar, M.S., Routh, A.L., Ren, P., Ku, Z., An, Z., Debbink, K., Diamond, M.S., Shi, P.-Y., Freiberg, A.N., Menachery, V.D., 2021. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. Nature 591, 293–299. https://doi.org/10.1038/s41586-021-03237-4

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

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285

Kimura, I., Yamasoba, D., Nasser, H., Zahradnik, J., Kosugi, Y., Wu, J., Nagata, K., Uriu, K., Tanaka, Y.L., Ito, J., Shimizu, R., Tan, T.S., Butlertanaka, E.P., Asakura, H., Sadamasu, K., Yoshimura, K., Ueno, T., Takaori-Kondo, A., Schreiber, G., Genotype to Phenotype Japan (G2P-Japan) Consortium, Toyoda, M., Shirakawa, K., Irie, T., Saito, A., Nakagawa, S., Ikeda, T., Sato, K., 2022. The SARS-CoV-2 spike S375F

mutation characterizes the Omicron BA.1 variant. iScience 25, 105720. https://doi.org/10.1016/j.isci.2022.105720
Lahti, D.C., Johnson, N.A., Ajie, B.C., Otto, S.P., Hendry, A.P., Blumstein, D.T., Coss, R.G., Donohue, K., Foster, S.A., 2009. Relaxed selection in the wild. Trends in Ecology & Evolution 24, 487–496. https://doi.org/10.1016/j.tree.2009.03.010
Lamers, M.M., Mykytyn, A.Z., Breugem, T.I., Groen, N., Knoops, K., Schipper, D., Acker, R. van, Doel, P.B. van den, Bestebroer, T., Koopman, C.D., Reusken, C., Muraro, M.J., GeurtsvanKessel, C.H., Royen, M.E. van, Peters, P.J., Zhang, J., Haagmans, B.L., 2022. SARS-CoV-2 Omicron efficiently infects human airway, but not alveolar epithelium. bioRxiv 2022.01.19.476898. https://doi.org/10.1101/2022.01.19.476898
Lavine, J.S., Bjornstad, O.N., Antia, R., 2021. Immunological characteristics govern the transition of COVID-19 to endemicity. Science 371, 741–745. https://doi.org/10.1126/science.abe6522
Leung, K., Shum, M.H., Leung, G.M., Lam, T.T., Wu, J.T., 2021. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Eurosurveillance 26, 2002106. https://doi.org/10.2807/1560-7917.ES.2020.26.1.2002106
Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv.
Liu, Y., Liu, J., Plante, K.S., Plante, J.A., Xie, X., Zhang, X., Ku, Z., An, Z., Scharton, D., Schindewolf, C., Widen, S.G., Menachery, V.D., Shi, P.-Y., Weaver, S.C., 2022. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. Nature 602, 294–299. https://doi.org/10.1038/s41586-021-04245-0
Lythgoe, K., Golubchik, T., Hall, M., House, T., Cahuantzi, R., MacIntyre-Cockett, G., Fryer, H., Thomson, L., Nurtay, A., Ghafani, M., Buck, D., Green, A., Trebes, A., Piazza, P., Lonie, L.J., Studley, R., Rourke, E., Smith, D., Bashton, M., Nelson, A., Crown, M., McCann, C., Young, G.R., Santos, R.A.N. dos, Richards, Z., Tariq, A., Team, W.S.I.C.-19 S., Group, C.-19 I.S., Consortium, T.C.-19 G.U. (COG-U., Fraser, C., Diamond, I., Barrett, J., Walker, A.S., Bonsall, D., 2023. Lineage replacement and evolution captured by three years of the United Kingdom Covid Infection Survey. https://doi.org/10.1101/2022.01.05.21268323
Markov, P.V., Ghafari, M., Beer, M., Lythgoe, K., Simmonds, P., Stilianakis, N.I., Katzourakis, A., 2023. The evolution of SARS-CoV-2. Nat Rev Microbiol 21, 361–379. https://doi.org/10.1038/s41579-023-00878-2
Minor, P.D., 2015. Live attenuated vaccines: Historical successes and current challenges. Virology, 60th Anniversary Issue 479–480, 379–392. https://doi.org/10.1016/j.virol.2015.03.032
Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., Köster, J., 2021. Sustainable data analysis with Snakemake. F1000Res 10, 33. https://doi.org/10.12688/f1000research.29032.2
Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32, 268–274. https://doi.org/10.1093/molbev/msu300
Ogando, N.S., Dalebout, T.J., Zevenhoven-Dobbe, J.C., Limpens, R.W.A.L., van der Meer, Y., Caly, L., Druce, J., de Vries, J.J.C., Kikkert, M., Bárcena, M., Sidorov, I., Snijder, E.J., 2020. SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. J Gen Virol 101, 925–940. https://doi.org/10.1099/jgv.0.001453

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J.T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., Du Plessis, L., Maloney, D., Medd, N., Attwood, S.W., Aanensen, D.M., Holmes, E.C., Pybus, O.G., Rambaut, A., 2021. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. Virus Evolution. https://doi.org/10.1093/ve/veab064

Pedersen, B.S., Quinlan, A.R., 2017. cyvcf2: fast, flexible variant analysis with Python. Bioinformatics 33, 1867–1869. https://doi.org/10.1093/bioinformatics/btx057

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Redondo, N., Zaldívar-López, S., Garrido, J.J., Montoya, M., 2021. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. Frontiers in Immunology 12.

Saito, A., Irie, T., Suzuki, R., Maemura, T., Nasser, H., Uriu, K., Kosugi, Y., Shirakawa, K., Sadamasu, K., Kimura, I., Ito, J., Wu, J., Iwatsuki-Horimoto, K., Ito, M., Yamayoshi, S., Loeber, S., Tsuda, M., Wang, L., Ozono, S., Butlertanaka, E.P., Tanaka, Y.L., Shimizu, R., Shimizu, K., Yoshimatsu, K., Kawabata, R., Sakaguchi, T., Tokunaga, K., Yoshida, I., Asakura, H., Nagashima, M., Kazuma, Y., Nomura, R., Horisawa, Y., Yoshimura, K., Takaori-Kondo, A., Imai, M., Tanaka, S., Nakagawa, S., Ikeda, T., Fukuhara, T., Kawaoka, Y., Sato, K., 2022. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. Nature 602, 300–306. https://doi.org/10.1038/s41586-021-04266-9

Sasaki, M., Uemura, K., Sato, A., Toba, S., Sanaki, T., Maenaka, K., Hall, W.W., Orba, Y., Sawa, H., 2021. SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells. PLOS Pathogens 17, e1009233. https://doi.org/10.1371/journal.ppat.1009233

Shuai, L., Zhong, G., Yuan, Q., Wen, Z., Wang, C., He, X., Liu, R., Wang, J., Zhao, Q., Liu, Y., Huo, N., Deng, J., Bai, J., Wu, H., Guan, Y., Shi, J., Tian, K., Xia, N., Chen, H., Bu, Z., 2020. Replication, pathogenicity, and transmission of SARS-CoV-2 in minks. Natl Sci Rev 8, nwaa291. https://doi.org/10.1093/nsr/nwaa291

Takada, K., Ueda, M.T., Shichinohe, S., Kida, Y., Ono, C., Matsuura, Y., Watanabe, T., Nakagawa, S., 2023. Genomic diversity of SARS-CoV-2 can be accelerated by mutations in the nsp14 gene. iScience 26, 106210. https://doi.org/10.1016/j.isci.2023.106210

Tay, J.H., Porter, A.F., Wirth, W., Duchene, S., 2022. The Emergence of SARS-CoV-2 Variants of Concern Is Driven by Acceleration of the Substitution Rate. Molecular Biology and Evolution 39, msac013. https://doi.org/10.1093/molbev/msac013

Viana, R., Moyo, S., Amoako, D.G., Tegally, H., Scheepers, C., Althaus, C.L., Anyaneji, U.J., Bester, P.A., Boni, M.F., Chand, M., Choga, W.T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., Giovanetti, M., Hardie, D., Hill, V., Hsiao, N.-Y., Iranzadeh, A., Ismail, A., Joseph, C., Joseph, R., Koopile, L., Kosakovsky Pond, S.L., Kraemer, M.U.G., Kuate-Lere, L., Laguda-Akingba, O., Lesetedi-Mafoko, O., Lessells, R.J., Lockman, S., Lucaci, A.G., Maharaj, A., Mahlangu, B., Maponga, T., Mahlakwane, K., Makatini, Z., Marais, G., Maruapula, D., Masupu, K., Matshaba, M., Mayaphi, S., Mbhele, N., Mbulawa, M.B., Mendes, A., Mlisana, K., Mnguni, A., Mohale, T., Moir, M., Moruisi, K., Mosepele, M., Motsatsi, G., Motswaledi, M.S., Mphoyakgosi, T., Msomi, N., Mwangi, P.N., Naidoo, Y., Ntuli, N., Nyaga, M., Olubayo, L., Pillay, S., Radibe, B., Ramphal, Y., Ramphal, U., San, J.E., Scott, L., Shapiro, R., Singh, L., Smith-Lawrence, P., Stevens, W., Strydom, A., Subramoney, K., Tebeila, N., Tshiabuila, D., Tsui, J., van Wyk, S., Weaver, S., Wibmer, C.K., Wilkinson, E., Wolter, N., Zarebski, A.E., Zuze, B., Goedhals, D., Preiser, W., Treurnicht, F., Venter, M., Williamson, C., Pybus, O.G., Bhiman, J., Glass, A., Martin, D.P., Rambaut, A., Gaseitsiwe, S., von

Gottberg, A., de Oliveira, T., 2022. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature 603, 679–686. https://doi.org/10.1038/s41586-022-04411-y

Weigang, S., Fuchs, J., Zimmer, G., Schnepf, D., Kern, L., Beer, J., Luxenburger, H., Ankerhold, J., Falcone, V., Kemming, J., Hofmann, M., Thimme, R., Neumann-Haefelin, C., Ulferts, S., Grosse, R., Hornuss, D., Tanriver, Y., Rieg, S., Wagner, D., Huzly, D., Schwemmle, M., Panning, M., Kochs, G., 2021. Within-host evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient as a source of immune escape variants. Nat Commun 12, 6405. https://doi.org/10.1038/s41467-021-26602-3

Yamamoto, M., Tomita, K., Hirayama, Y., Inoue, J., Kawaguchi, Y., Gohda, J., 2022. SARS-CoV-2 Omicron spike H655Y mutation is responsible for enhancement of the endosomal entry pathway and reduction of cell surface entry pathways. https://doi.org/10.1101/2022.03.21.485084

Yuan, S., Ye, Z.-W., Liang, R., Tang, K., Zhang, A.J., Lu, G., Ong, C.P., Man Poon, V.K., Chan, C.C.-S., Mok, B.W.-Y., Qin, Z., Xie, Y., Chu, A.W.-H., Chan, W.-M., Ip, J.D., Sun, H., Tsang, J.O.-L., Yuen, T.T.-T., Chik, K.K.-H., Chan, C.C.-Y., Cai, J.-P., Luo, C., Lu, L., Yip, C.C.-Y., Chu, H., To, K.K.-W., Chen, H., Jin, D.-Y., Yuen, K.-Y., Chan, J.F.-W., 2022. Pathogenicity, transmissibility, and fitness of SARS-CoV-2 Omicron in Syrian hamsters. Science 377, 428–433. https://doi.org/10.1126/science.abn8939

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S.B., Schaffner, S.F., Lemieux, J.E., Munro, J.B., Rafique, A., Barve, A., Sabeti, P.C., Kyratsous, C.A., Dudkina, N.V., Shen, K., Luban, J., 2020. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell 183, 739-751.e8. https://doi.org/10.1016/j.cell.2020.09.032

Zandi, M., Shafaati, M., Kalantar-Neyestanaki, D., Pourghadamyari, H., Fani, M., Soltani, S., Kaleji, H., Abbasi, S., 2022. The role of SARS-CoV-2 accessory proteins in immune evasion. Biomedicine & Pharmacotherapy 156, 113889. https://doi.org/10.1016/j.biopha.2022.113889

Zhou, B., Thao, T.T.N., Hoffmann, D., Taddeo, A., Ebert, N., Labroussaa, F., Pohlmann, A., King, J., Steiner, S., Kelly, J.N., Portmann, J., Halwe, N.J., Ulrich, L., Trüeb, B.S., Fan, X., Hoffmann, B., Wang, L., Thomann, L., Lin, X., Stalder, H., Pozzi, B., de Brot, S., Jiang, N., Cui, D., Hossain, J., Wilson, M.M., Keller, M.W., Stark, T.J., Barnes, J.R., Dijkman, R., Jores, J., Benarafa, C., Wentworth, D.E., Thiel, V., Beer, M., 2021. SARS-CoV-2 spike D614G change enhances replication and transmission. Nature 592, 122–127. https://doi.org/10.1038/s41586-021-03361-1

Zinzula, L., 2021. Lost in deletion: The enigmatic ORF8 protein of SARS-CoV-2. Biochemical and Biophysical Research Communications, COVID-19 538, 116–124. https://doi.org/10.1016/j.bbrc.2020.10.045

Figures

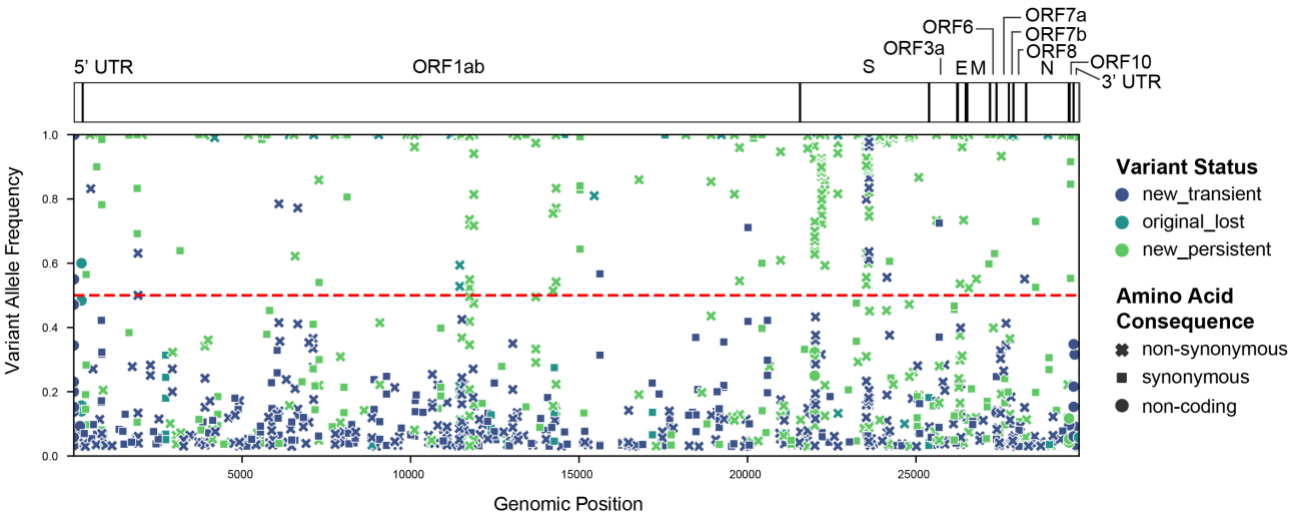


Figure 1: The location of all detected variants across all samples in all passage lines. The colour of points indicates whether a variant was present in the original clinical isolate ('original_lost'), or arose new during serial passaging but was then lost ('new_transient') or persisted in the final passage number ('new_persistent'). The shape of points corresponds to the inferred amino acid consequence, either non-synonymous, synonymous, or 'other' (in non-coding regions). All points above the horizontal dashed line are the majority allele for a given sample (allele frequency ≥50%). The location of genes within the SARS-CoV-2 genome is indicated above the plot in a schematic of the complete SARS CoV-2 genome of ~30 kb.

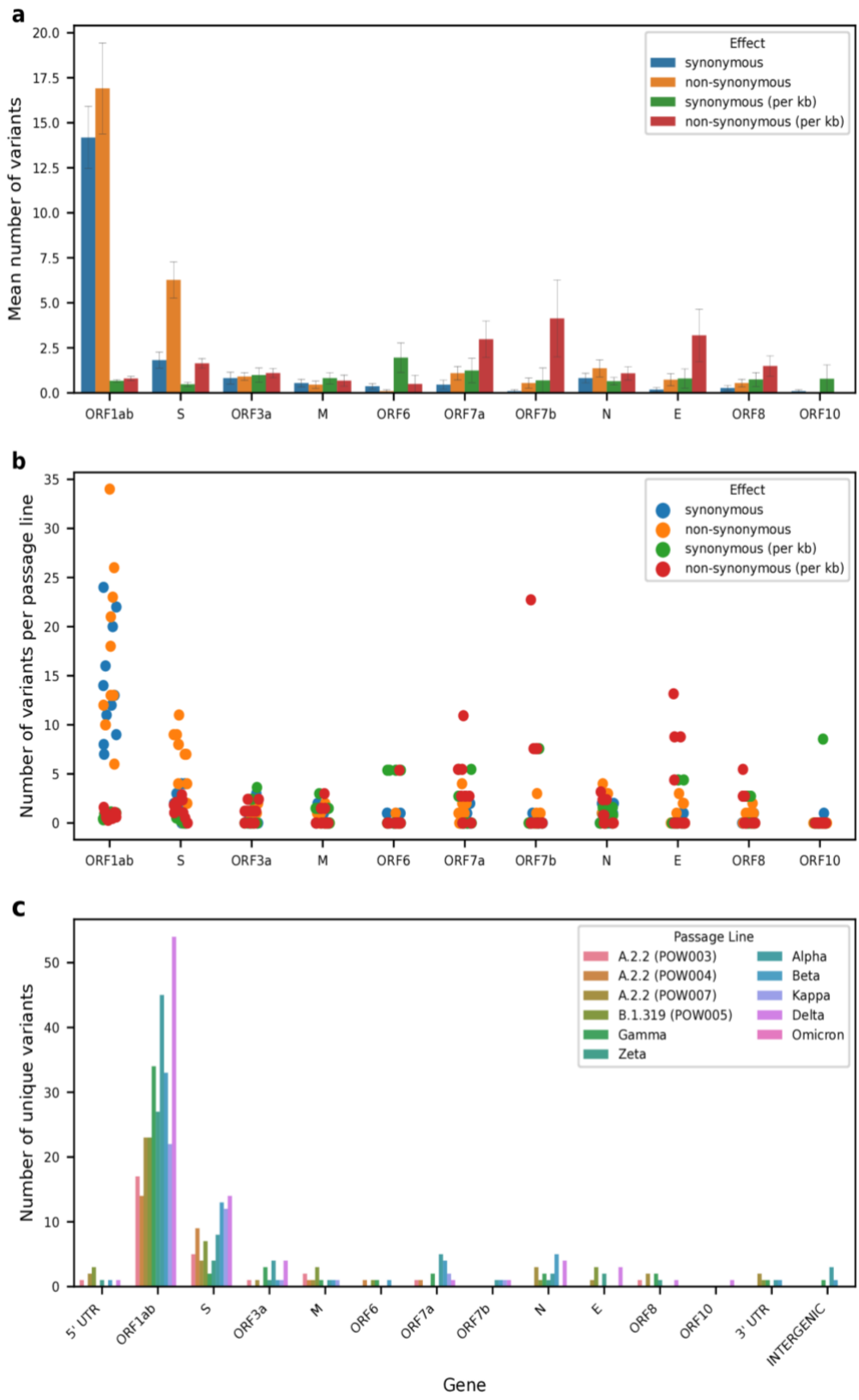


Figure 2: The location of new variants arising during serial passaging within the SARS-CoV-2 genome. **a** Bars represent the mean number among passage lines (error bars: 95% confidence intervals), with the shading representing the inferred amino acid consequences of the variants both with and without scaling by gene length. **b** As per panel 'a', but with each point representing a sample within a given passage line. **c** The number of unique variants detected per genomic region, separated by passage line.

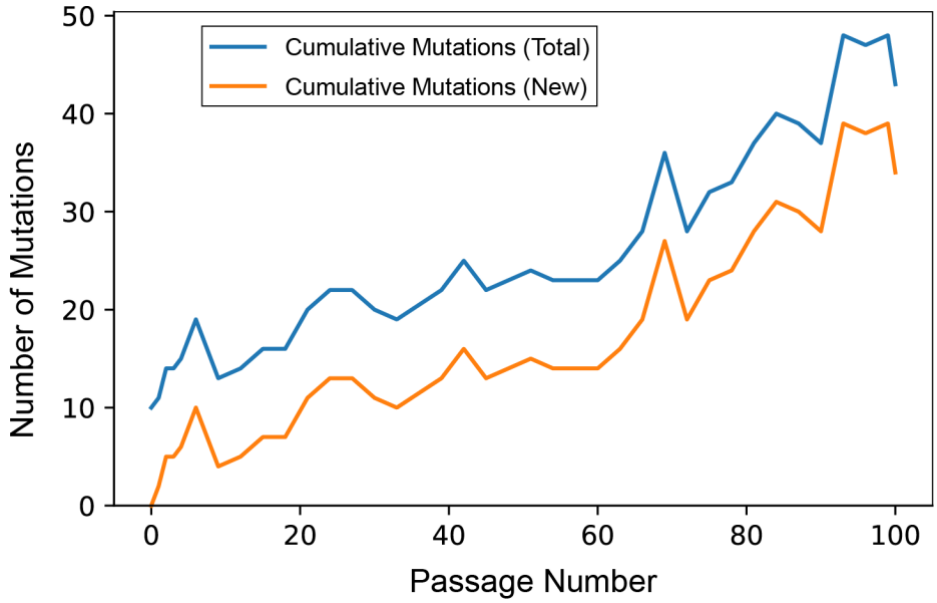


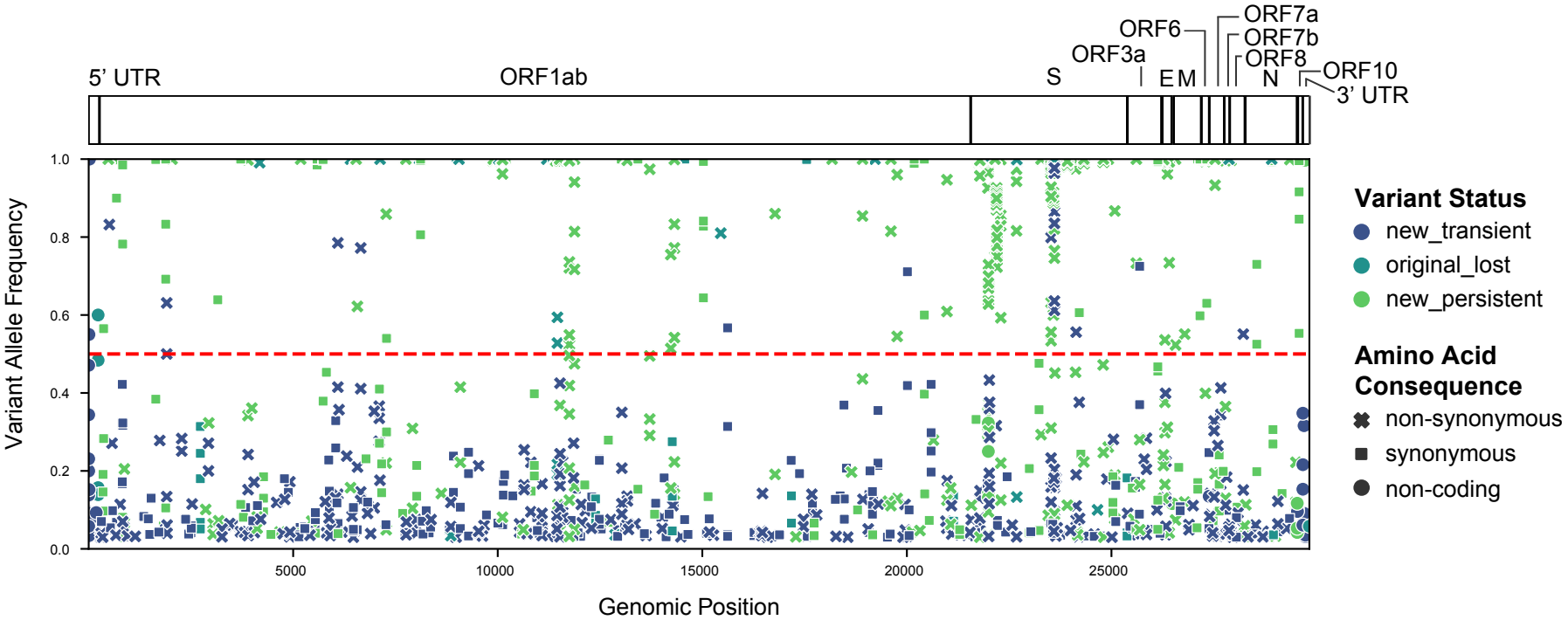
Figure 3: The cumulative number of mutations detected within passage line “POW005” (Pango lineage: B.1.319) over 100 serial passages. The number of mutations at any given passage number refers to those mutations detected with respect to the SARS-CoV-2 reference genome. The total number is represented by the blue line, with the orange line representing only those mutations that newly arose throughout serial passaging (absent from the clinical isolate).

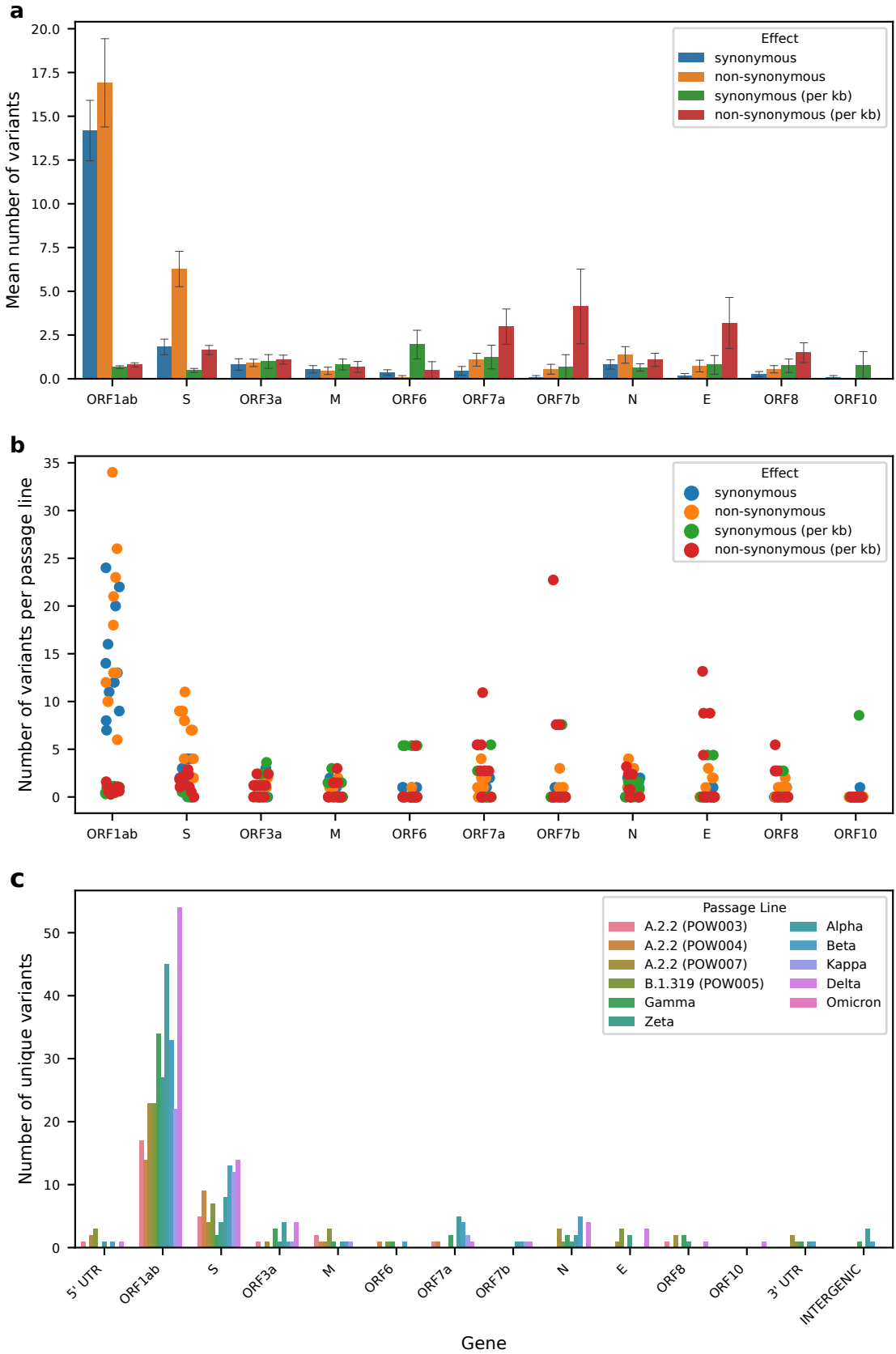
Tables

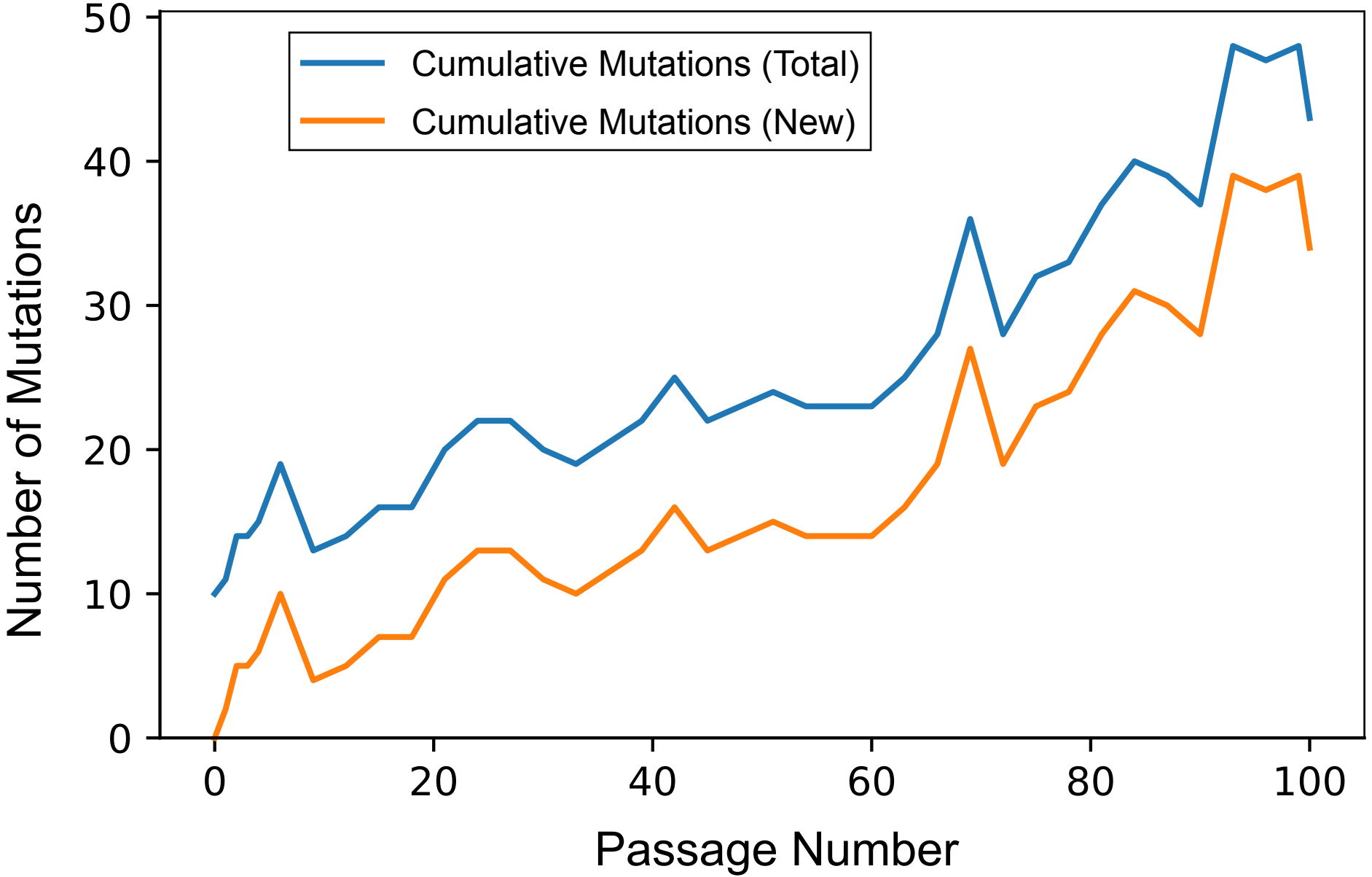
Table 2: All mutations arising *de novo* throughout the course of long-term serial passaging that are noted in a compendium of functional impacts of mutations in the SARS-CoV-2 Wuhan Hu-1 reference genome (https://github.com/nodrogluap/pokay). All passage lines we studied (see text) are noted in the first column. Reference numbers refer to those in Supplementary File S1.

Passage Line	Gene	Amino Acid Consequence	Category	Reference(s)
A.2.2 (POW003)	M	H125Y	frequency based fitness	1
A.2.2 (POW003)	M	H125Y	homoplasy	2
A.2.2 (POW004)	S	S686G	anthropozoonotic events	3
A.2.2 (POW007)	E	4F	frequency based fitness	1
Alpha	S	S813I	convalescent plasma escape	4; 5
B.1.319 (POW005)	S	F338L	gene expression increase	6
B.1.319 (POW005)	S	F338L	monoclonal antibody serial passage escape	7
B.1.319 (POW005), A.2.2 (POW004)	M	53F	frequency based fitness	1
B.1.319 (POW005), Alpha	S	N74K	homoplasy	2
B.1.319 (POW005), Zeta	E	T30I	homoplasy	2
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	ACE2 receptor binding affinity	8; 9; 10; 11
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	T cell evasion	12; 13
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	anthropozoonotic events	14; 15; 16; 17
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	antibody epitope effects	18; 19; 20; 21; 22; 23; 24; 25; 26
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	convalescent plasma binding	9; 11
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	convalescent plasma escape	18; 27; 28; 29; 30; 22; 31; 25; 32; 33; 34; 35
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	homoplasy	2
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	outcome hazard ratio	36; 37; 38
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	pharmaceutical effectiveness	27; 39; 40; 41; 42; 34; 43
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	reinfection	44; 45
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	trafficking	17; 34
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	transmissibility	46; 47
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	vaccine efficacy	48; 49
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	vaccine neutralization efficacy	50; 51; 13; 10; 52; 26; 35; 43; 53; 54; 55; 11; 9; 56; 32; 18; 12; 57; 27; 58; 30; 41; 21; 42; 25; 34; 59; 8; 22; 60; 61
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	vaccinee plasma binding	9; 11
Beta, A.2.2 (POW003), B.1.319 (POW005), Delta, A.2.2 (POW004), Zeta	S	H655Y	virion structure	18; 62; 22
Beta, A.2.2 (POW007), A.2.2 (POW004), Alpha	S	R682W	anthropozoonotic events	17

Beta, A.2.2 (POW007), A.2.2 (POW004), Alpha	S	R682W	vaccine neutralization efficacy	63; 64; 60; 65; 56; 66
Beta, Omicron (BA.1)	S	A67V	ACE2 receptor binding affinity	8; 10
Beta, Omicron (BA.1)	S	A67V	antibody epitope effects	20; 21; 23; 24; 26; 19
Beta, Omicron (BA.1)	S	A67V	pharmaceutical effectiveness	39; 40; 41; 42; 43
Beta, Omicron (BA.1)	S	A67V	vaccine neutralization efficacy	8; 58; 41; 21; 42; 61; 10; 52; 43; 26
Delta	S	A222V	antibody epitope effects	67
Delta	S	A222V	convalescent plasma escape	68; 69
Delta	S	A222V	humoral response durability	70
Delta	S	A222V	reinfection	70
Delta	S	A222V	vaccine neutralization efficacy	68; 69
Delta	S	T76I	antibody epitope effects	71
Delta	S	T76I	convalescent plasma escape	30; 32
Delta	S	T76I	vaccine neutralization efficacy	30; 32
Delta, Omicron (BA.1)	S	L5F	ACE2 receptor binding affinity	9; 11
Delta, Omicron (BA.1)	S	L5F	antibody epitope effects	72; 73
Delta, Omicron (BA.1)	S	L5F	convalescent plasma binding	9; 11
Delta, Omicron (BA.1)	S	L5F	convalescent plasma escape	74; 59; 75
Delta, Omicron (BA.1)	S	L5F	vaccine efficacy	48; 49
Delta, Omicron (BA.1)	S	L5F	vaccine neutralization efficacy	75; 54; 58; 11; 74; 9
Delta, Omicron (BA.1)	S	L5F	vaccinee plasma binding	9; 11
Gamma	S	43F	monoclonal antibody serial passage escape	76
Gamma	S	43F	pharmaceutical effectiveness	77; 78
Kappa	M	41N	frequency based fitness	1
Kappa	S	S704L	antibody epitope effects	26; 21
Kappa	S	S704L	vaccine neutralization efficacy	26; 21
Omicron (BA.1)	S	S373P	ACE2 receptor binding affinity	8; 79; 10; 80
Omicron (BA.1)	S	S373P	antibody epitope effects	20; 21; 23; 24; 26; 19
Omicron (BA.1)	S	S373P	monoclonal antibody serial passage escape	7
Omicron (BA.1)	S	S373P	pharmaceutical effectiveness	43; 41; 42; 40
Omicron (BA.1)	S	S373P	transmissibility	81
Omicron (BA.1)	S	S373P	vaccine neutralization efficacy	8; 41; 21; 61; 10; 52; 26
Zeta	E	T9I	frequency based fitness	1
Zeta	E	T9I	homoplasy	2


5' UTR
ORF1ab
S
ORF3a
E
M
ORF6
ORF7a
ORF7b
ORF8
N
ORF10
3' UTR
Variant Allele Frequency
1.0
0.8
0.6
0.4
0.2
0.0
5000
10000
15000
20000
25000
Genomic Position
Variant Status
new_transient
original_lost
new_persistent
Amino Acid Consequence
non-synonymous
synonymous
non-coding


a
Mean number of variants
Effect
synonymous
non-synonymous
synonymous (per kb)
non-synonymous (per kb)
ORF1ab
S
ORF3a
M
ORF6
ORF7a
ORF7b
N
E
ORF8
ORF10
b
Number of variants per passage line
Effect
synonymous
non-synonymous
synonymous (per kb)
non-synonymous (per kb)
c
Number of unique variants
Passage Line
A.2.2 (POW003)
A.2.2 (POW004)
A.2.2 (POW007)
B.1.319 (POW005)
Gamma
Zeta
Alpha
Beta
Kappa
Delta
Omicron
5' UTR
INTERGENIC
3' UTR
Gene


Cumulative Mutations (Total)
Cumulative Mutations (New)
Number of Mutations
Passage Number
0
10
20
30
40
50
0
20
40
60
80
100