

## 2022: Immunobiology and Pathogenesis of SARS-CoV-2 infection

### COVID-19, the first pandemic in the post-genomic era

Lucy van Dorp<sup>1</sup>, Charlotte J. Houldcroft<sup>2</sup>, Damien Richard<sup>1,3</sup>, François Balloux<sup>1</sup>

<sup>1</sup>UCL Genetics Institute, University College London, London, WC1E 6BT, UK.

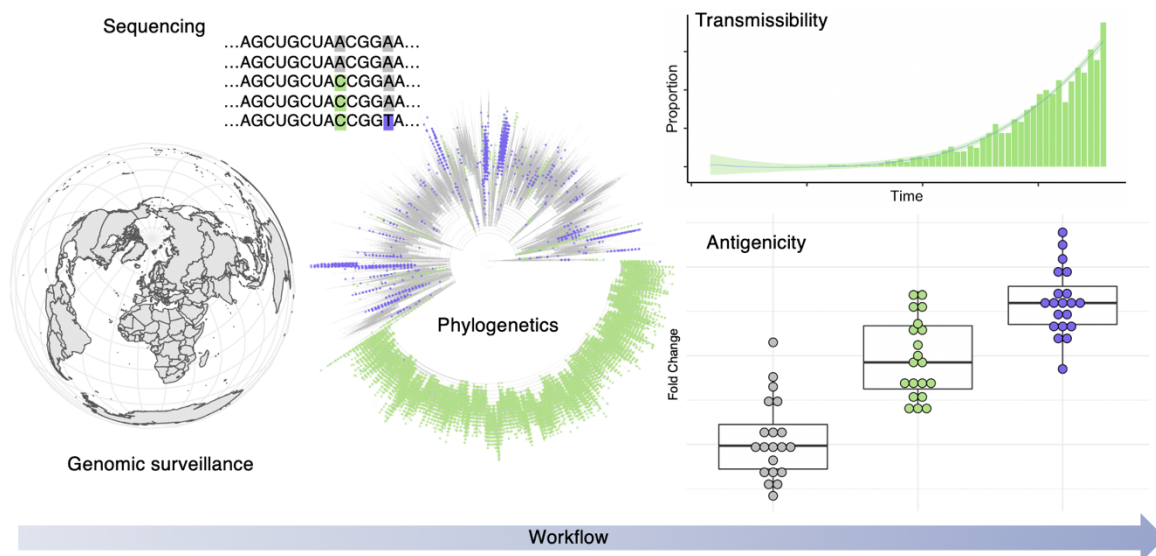
<sup>2</sup>Department of Medicine, University of Cambridge, Cambridge, CB2 1TN, UK.

<sup>3</sup>Division of Infection and Immunity, University College London, London, WC1E 6BT, UK.

#### Abstract [128 words]

The scale of the international efforts to sequence SARS-CoV-2 genomes is unprecedented. Early availability of genomes allowed rapid characterisation of the virus, thus kickstarting many highly successful vaccine development programmes. Worldwide genomic resources have provided a good understanding of the pandemic, supported close monitoring of the emergence of viral genomic diversity and pinpointed those sites to prioritise for functional characterisation. Continued genomic surveillance of global viral populations will be crucial to inform the timing of vaccine updates so as to pre-empt the spread of immune escape lineages. While genome sequencing has provided us with an exceptionally powerful tool to monitor the evolution of SARS-CoV-2, there is room for further improvements in particular in the form of less heterogeneous global surveillance and tools to rapidly identify concerning viral lineages.

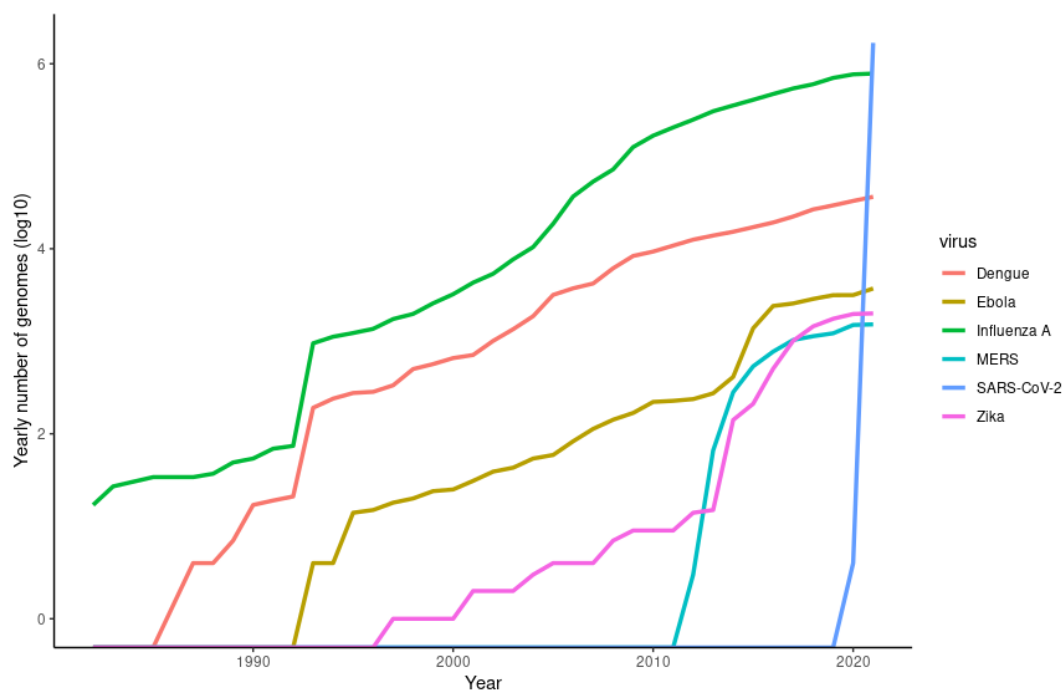
#### Graphical Abstract



## Introduction

The COVID-19 pandemic has shone a spotlight on the value of large-scale, open, and near real-time genomic surveillance of pathogens. The first whole genome sequences of Severe Acute Respiratory Coronavirus 2 (SARS-CoV-2) were available within ten days after a cluster of cases of pneumonia in Wuhan, China was reported to the World Health Organisation (WHO) on the 31<sup>st</sup> December 2019. Eighteen months later, close to two million SARS-CoV-2 genomes have been made available. This resource offers unprecedented opportunities to monitor the emergence of viral genomic diversity and to reconstruct transmission routes as SARS-CoV-2 continues to adapt to its human host.

COVID-19 is arguably the first post-genomic pandemic. By the time the WHO officially declared the pandemic on the 11<sup>th</sup> of March 2020 over 500 whole genome sequences had been shared spanning 39 countries and six continents. To place this in perspective, an analysis published shortly after the declaration of the H1N1pdm influenza pandemic in late April 2009, comprised only 11 partial hemagglutinin (HA) sequences<sup>1</sup>. More recently, close to real time surveillance efforts have come to the fore in the reconstruction of transmission of Zika and Ebola, aided by portable sequencing devices<sup>2,3</sup>, though these remain relatively limited in the number of genomes sequenced. Influenza genomic surveillance, which represents a cornerstone of the biannual assessment of the multivalent flu vaccines, has led to the generation of close to a million genome sequences over the last 30 years (**Figure 1**). Though, this number has been overtaken by the ongoing sequencing effort for SARS-CoV-2 within 18 months with 2 million genomes available by the end of June 2021, and rising. Active monitoring of viral evolution has become, and will likely remain, a mainstay of pandemic response, both for COVID-19 and for future epidemics.



**Figure 1:** Cumulative number of genome submissions to NCBI virus for major human viral pathogens for which genomic surveillance has been deployed. Data for SARS-CoV-2 is obtained from the GISAID data sharing repository. Y-axis provides the  $\log_{10}$ (cumulative number of genomes) and the x-axis provides the date of sample collection spanning from 1982 until the time of writing (June 2021). Lines are smoothed based on observations per year.

## Tracking the emergence of genomic diversity

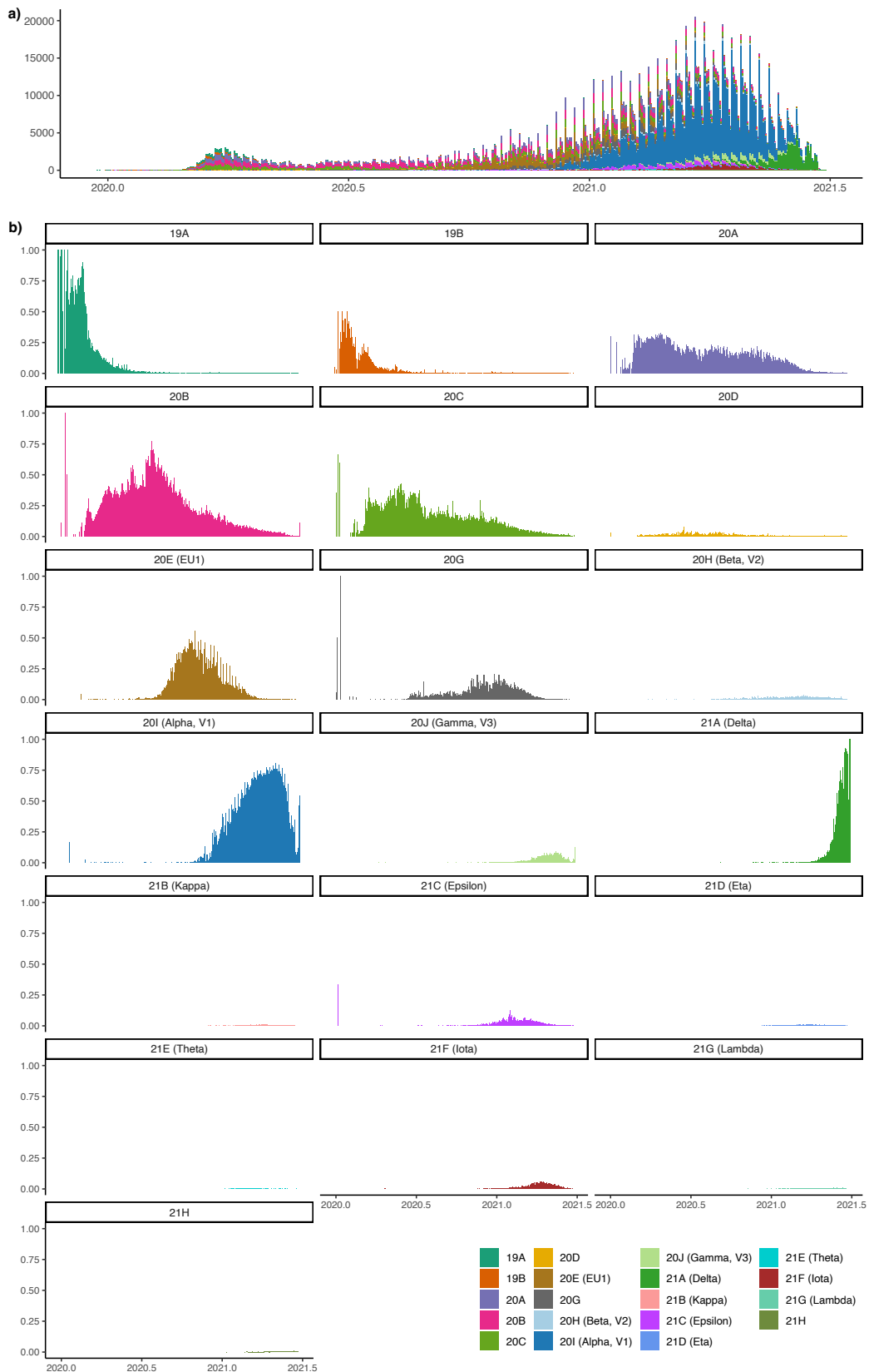
### *Early observations*

The rapid description of the first SARS-CoV-2 genome on the 10<sup>th</sup> of January 2020 was vital to identify the previously unknown coronavirus SARS-CoV-2 and represented the first step towards vaccine development and genome sequencing initiatives<sup>4,5</sup>. Genomes uploaded over the first few months of 2020 facilitated initial assessments of the evolutionary rate of SARS-CoV-2, of approximately ~2 mutations per month. This rate is suggestive of a most recent common ancestor of sampled pandemic lineages to the latter portion of 2019<sup>6-8</sup>, and a rapid spread to Europe, as confirmed by SARS-CoV-2 positive wastewater samples from Northern Italy dating to December 2019<sup>9</sup>.

### *Lineage dynamics*

Several studies early in the pandemic identified multiple independent introductions into regions of the world during early 2020<sup>10-14</sup>. For example, the first epidemic wave in the UK was seeded by well over 1000 independent SARS-CoV-2 introductions<sup>10</sup>. As a result of this extensive global spread, the worldwide SARS-CoV-2 population first remained geographically largely unstructured with the same major cosmopolitan lineages found in most regions of the world. The introduction of travel-bans and regional restrictions in early 2020 led to the emergence of more geographically associated lineages, though not precluding the capacity for introduced lineages to have marked impacts on local lineage frequencies<sup>15</sup>. Ongoing tracing of SARS-CoV-2 over the course of the COVID-19 pandemic has highlighted striking dynamics (**Figure 2**), with major clades changing markedly in frequency through time.

At the finer-scale, identification of SARS-CoV-2 lineages supports reconstruction of local chains of transmission though can be complicated by factors such as inter-individual variability in the SARS-CoV-2 incubation periods, asymptomatic cases, and missing transmission nodes. Reconstruction of transmission chains can be challenging in the presence of high community infection levels and/or multiple introduction events. Nonetheless, genomic analysis of SARS-CoV-2 outbreaks has helped to identify unanticipated sources of transmission and to rule out transmission events suspected on the basis of e.g. traditional contact tracing or movement data<sup>13,16,17</sup>.



**Figure 2:** a) Daily counts of genome assemblies (y-axis) shared on GISAID (current to 5/7/2021) over the course of the pandemic for all NextStrain clades. Note 20I (Alpha, V1) is broadly equivalent to PANGO lineage B.1.1.7, 20H (Beta, V2) to B.1.351, 20J (Gamma, V3) to P.1. and 21A (Delta) to B.1.617.2. b) Equivalent plots as a) split on each SARS-CoV-2 clade providing the daily proportions (y-axis) of genome submissions. Colour assignments for each clade are given as per the legend at bottom right.

### *Genomics Infrastructure*

GISAID<sup>18,19</sup> is an instrumental platform for SARS-CoV-2 genome sharing, supporting appropriate accreditation of data providers, a tenet central to the data sharing model. This resource has supported several initiatives allowing visualisation and mutation characterisation including NextStrain<sup>20</sup>, CoVizu<sup>21</sup>, cov-lineages.org<sup>22</sup>, outbreak.info (<https://outbreak.info/>) and CoVariants (<https://covariants.org>). In addition, large-scale assessment of phylogenies is greatly aided by the GISAID Audacity workflow which provides readily updated phylogenies across high-quality genome submissions. However, there is a growing shortfall in methods that can tractably keep pace with expanding datasets (**Figure 1**), potentially posing a major bottleneck to continued efforts to track diversity<sup>23</sup>. In part this is because standard phylogenetic pipelines were not designed for datasets as large, intensively sampled and with the low genetic diversity characteristic of SARS-CoV-2 genomic studies. One solution is dataset subsampling, which is a suitable approach for certain questions, but not fully satisfactory for mutation tracking efforts. In particular, recurrent convergent mutations are an important feature of concerning variants (**Figure 3**), hence considering all independent emergences over large-scale phylogenies is vital to assess the impact of mutations in different spatiotemporal contexts<sup>24,25</sup>.

Methodological advances to circumvent these challenges have included rapid maximum parsimony placement of new genomes onto existing tree topologies<sup>26</sup> and reconstruction of mutational history bypassing phylogenetic reconstruction<sup>8</sup>. Additionally, there has been a need to devise a nomenclature to characterise emerging diversity, with the PANGO lineage assignment tool having become a mainstay in genomic epidemiology studies of SARS-CoV-2<sup>22</sup>. Other efforts have focused on defining wider phylogenetic ‘clades’ with both GISAID<sup>18,19</sup> and NextStrain<sup>20</sup> providing schemes (**Figure 2**). The WHO have also established a nomenclature system, based around letters of the Greek alphabet, to aid public discussion of concerning variants (<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>, accessed 5/7/2021).

### **Adaptive changes**

#### *Initial adaptive mutations*

Following a host jump into a new species, a viral pathogen must adapt to the host cell machinery and avoid immune defences to successfully persist and transmit. Over the first nine months of the pandemic, the SARS-CoV-2 population remained reasonably evolutionarily stable with the exception of the emergence of spike D614G in February 2020 and accompanying sites (nucleotides 241, 3037, 14408)<sup>27</sup>. Lineages carrying 614G now represent the majority of sequence data (**Figure 2**), with this dominance likely also aided by founder effects during the early pandemic spread<sup>28</sup>. Further early described adaptive changes include N439K, located in the spike Receptor Binding Domain (RBD), which confers partial resistance to several neutralizing monoclonal antibodies and enhances binding to human ACE2 receptors<sup>29</sup>. Spike mutation A222V, largely associated with PANGO lineage B.1.177 (clade 20E EU1), also rose markedly in frequency over the Northern hemisphere summer of 2020. However its adaptive potential is debated, with its rapid growth in Europe suggested to be due to seeding by travel associated infections rather than the mutation conferring an intrinsic transmission advantage<sup>15</sup> (**Figure 2**).

Insights into the early stages of adaptation following a host switch were also provided by subsequent SARS-CoV-2 host jumps into farmed minks<sup>30</sup>. Viral lineages circulating in mink independently acquired a set of mutations including spike Y453F<sup>30,31</sup>, which is rarely

observed in humans (**Figure 3**). The combination of Y453F, spike del69/70, I692V and M1229I defined ‘cluster 5’, a lineage of SARS-CoV-2 which exhibited reduced antibody neutralization<sup>32,33</sup>, and jumped back into humans before going extinct. Though, the role of mink farms as animal reservoirs potentially fuelling transmission in the human population prompted authorities in Denmark to proceed with mass culling of minks.

#### *Emergence of Variants of Concern*

In late 2020 three distinct SARS-CoV-2 lineages emerged almost concurrently in different continents. All have spike mutation N501Y and came to attention due to their unusual combination and number of mutations, together with their detection coinciding with rapidly increasing cases in regions where they were first linked to (**Figure 3**). These include Alpha (B.1.1.7, clade 20I/501Y.v1) first detected in the UK<sup>34</sup>, Beta (B.1.351, 20H/501Y.v2) in South Africa<sup>35</sup>, and Gamma (P.1, 20J/501Y.v3) first identified in cases linked to Brazil<sup>36</sup>. Each lineage was elevated to the status of a Variant of Concern (VoC), a term employed by public health agencies to define a lineage with concerning epidemiological, immunological, or pathogenic properties. These VoCs (henceforth 501Y VoCs) were demonstrated to show enhanced receptor binding<sup>37–39</sup> and increased transmissibility to varying degrees<sup>36,40</sup> with B.1.1.7 the most transmissible 501Y VoC in circulation<sup>25</sup>. Beta and Gamma additionally demonstrated some ability to bypass immunity elicited by prior infection or vaccination<sup>33,41–45</sup> (**Box 1**). In addition, both Alpha and Gamma have been associated with more virulent infections in some settings<sup>46–48</sup>. Around the same time period, two lineages were identified in California (B.1.427 and B.1.429; CAL.20C) coinciding with a period of rapid pandemic growth<sup>49</sup>. These lineages carry another RBD mutation L452R which is also a hallmark of Delta (B.1.617.2)<sup>50</sup>, ascribed as a VOC in May 2021, which was linked to the epidemics in India during the first half of 2021 and then rapidly rose to dominance in the UK in May 2021 and subsequently in many countries throughout the world (**Figure 3**).

#### *Candidates and mechanisms*

The appearance of similar mutations in diverse lineage backbones provides an indication of convergent evolution<sup>6</sup>, with studies of 501Y VoCs supporting a marked selective shift in the fitness of SARS-CoV-2<sup>25,51</sup> (**Figure 3**). N501Y is a highly recurrent mutation which exhibits increased binding affinity to human ACE2 receptors<sup>38</sup>. Modelling of protein dynamics suggest this mutation leads to a conformation shift in the spike to favour an open state<sup>52</sup>, aiding viral entry into human cells. Other major recurrent changes in the RBD include E484K which has also been suggested to increase receptor binding, particularly when in combination with N501Y<sup>37</sup>. E484K has raised additional concern as it may significantly reduce both convalescent and virus neutralisation by sera obtained from convalescent and vaccinated individuals<sup>53–55</sup> (**Box 1**). Similarly L452R confers resistance to monoclonal antibodies<sup>56</sup> and partial escape of HLA-A24-mediated cellular immunity, coupled with increased receptor binding relative to wildtype<sup>57</sup>. Of note, similar antigenic effects have been suggested for changes at codon 417 (417N and 417T)<sup>55,58</sup>, though these mutations are conversely associated to a decrease in ACE2 binding<sup>29,38</sup>. Thus, the relationship between receptor binding affinity and transmissibility, including the effects when mutations occur in combination, is likely to be complex.

Spike mutations falling outside of the RBD are also be of adaptive importance. Indeed, changes at codon 681 are suggested to enhance the efficiency of the furin cleavage separating the S1 and S2 subunits<sup>59</sup>. Additionally, the S1 N-terminal domain (NTD) shows a propensity for recurrent deletions<sup>60</sup>, some of which are found in current VoCs (Alpha:del69/70,Δ144, Beta:del241/243; Delta:del157/158). Deletions in the NTD may contribute to altered antibody

recognition<sup>61,62</sup> with genetic variation in this region also suggested as supporting conformational changes in the spike protein<sup>63,64</sup>, either permitting or compensating for other changes in the RBD, with which they are often associated.

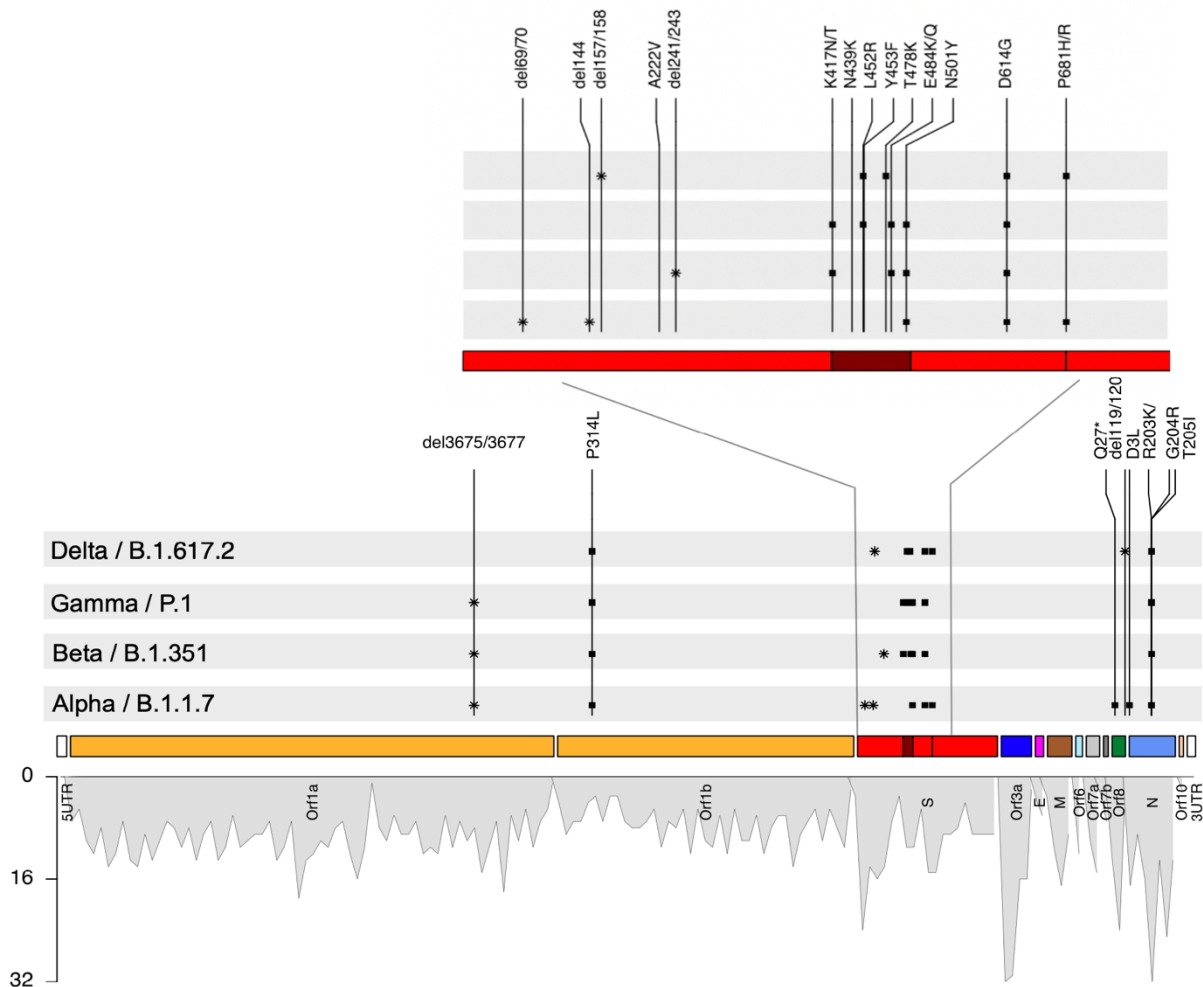
While changes in the spike protein remain major candidates in host adaptation, not least because of immuno-dominance and the role in cell entry, the genome-wide propensity for recurrent mutations in SARS-CoV-2 calls for a broader perspective (**Figure 3**). For instance, all aforementioned 501Y VoCs carry the same deletions at amino acid positions 106-108 of NSP6, a region thought to play an important role in innate immunity<sup>65</sup>. Other major shared changes are observed in the nucleocapsid protein, including widespread adjacent changes at codons 203-205. Also within the nucleocapsid, the D3L ‘triple mutation’ (deriving from 28280:GAT>CTA) has been demonstrated to enhance subgenomic RNA expression in Alpha<sup>66</sup> and computationally scored as enhancing transmissibility<sup>25</sup>. Outside of the structural proteins, deletions in ORF8 have also been implicated in altered pathogenicity<sup>67</sup>, with a truncated ORF8 observed in Alpha and deletions identified in Delta (**Figure 3**).

#### **Box 1: Quantifying the impact of viral evolution on host immunity**

There is concern over the extent to which previously SARS-CoV-2 exposed, or vaccinated individuals could be (re-)infected by emergent variants. High levels of humoral immunity (high antibody titres) are associated with protection from reinfection in animal models and studies of frequently re-exposed health-care workers<sup>68,69</sup>. Animal, human challenge or natural evidence of reinfection<sup>70</sup> provide more definitive answers on the ability of a lineage or clusters of mutations to evade the total (cellular and innate) immune response of the host; but these are slow, and may only provide useful data long after a new lineage has spread. As SARS-CoV-2 diversity increases, there is a need to rapidly test the putative immune escape phenotype of new viral mutants.

Neutralisation assays provide an *ex-vivo* method of screening for SARS-CoV-2 variants with the potential to escape humoral immune responses and can be performed with live or spike-pseudotyped virus, which have similar assay characteristics<sup>71</sup>. Following identification, new mutations within the spike protein can be engineered into a pseudotyped viral vector system or SARS-CoV-2 backbone. Pseudotyped spike mutants or viral isolates can then be incubated with heat-inactivated human serum at a range of serum dilutions and used to infect cells, with infectivity and viral dissemination quantified. Any variant with an increased ID50 value (i.e. more concentrated serum is required to neutralise the virus by at least 50%) can be said to show a degree of escape from humoral immunity.

Mutations within the RBD (including L452 and E484) are seen in a number of lineages associated with reduced vaccine-derived antibody neutralisation sensitivity<sup>72</sup>. For example, E484K on the Alpha variant background led to a six-fold decrease in neutralisation by serum from recipients of two BNT162b2 doses<sup>73</sup>. Similarly, neutralisation of live Beta-lineage virus was estimated to be 3-fold lower compared to wild-type in vaccinated individuals. Naturally infected, convalescent individuals in this study did not show such a pronounced drop in antibody neutralisation, which might reflect antibody responses to a broader range of epitopes (non-spike neutralising and non-neutralising) in the latter<sup>74</sup>.



**Figure 3:** Major mutations of interest along the SARS-CoV-2 genome. Line markers provide placement of mutations discussed in the text with the zoom box highlighting spike mutations. VoCs (defined current to May 2021) are highlighted for their carriage of major discussed mutations and deletions (\* deletion; ■ SNP), demonstrating striking patterns of convergent evolution. Note Public Health England additionally classify Alpha + E484K as a VoC. Major Variants of Interest (VoIs) at the time of writing include B.1.427 and B.1.429 (first detected in California and classified as VoCs by the Centers for Disease Control and Prevention; S:L452R). B.1.526 (first detected in New York; S:E484K + NSP6:del3675/3677), P.2 (first detected in Brazil; S:E484K), P.3 (first detected in the Philippines; S:E484K + S:N501Y), B.1.525 (detected in UK/Nigeria; NSP6:del3675/3677, S:del69/70, S:del144/145, S:E484K) and B.1.617.1 and B.1.617.3 (detected in India; S:452R, SE484Q, P681R). The bottom panel provides the density distribution of the count of recurrent mutations, over a 200-nucleotide sliding window, for the SARS-CoV-2 genome. The y-axis provides the number of estimated emergences based on a curated phylogeny of 550,743 sequences dating to April 2021<sup>25</sup>.

## Surveillance for future adaptive changes

### *Early detection*

Early detection of concerning lineages is critical for timely monitoring of viral transmission and deployment of appropriate interventions. This largely relies on the completeness of the picture of the SARS-CoV-2 population provided by genome sequencing. While some countries implemented close to real-time monitoring of SARS-CoV-2 in their surveillance protocols, this is not feasible in all settings. Alternatives to systematic patient-based sampling, for instance environmental monitoring of wastewater or sewage, may provide cost-effectiveness and a wider picture of the circulating genomic diversity in one given area<sup>75</sup>. Another possible strategy to circumvent the challenge of achieving representative surveillance is to focus on sequencing of incoming travellers and travel-associated infections



in high resource settings, as exemplified by Gamma which was identified in Japan in individuals with a history of travel to Brazil<sup>76</sup>.

Early detection of concerning lineages further relies on the ability to identify them from their genomic make-up. This is aided by flagging of mutations or combinations of mutations that have been demonstrated to have a phenotypic impact via functional studies (e.g. **Box 1**) or are suggested to be relevant using more unsupervised methods (e.g. scans of modified binding affinity<sup>38</sup> and/or mapping of escape mutations<sup>53</sup>). However, not all mutations have been assessed and such an approach ignores the possibility of epistatic (i.e. non-linear) effects. Due to this phenomenon, mutations that are neutral or even deleterious in isolation could provide a viral strain with a fitness increase when combined in the same background. Strong epistatic effects can lead to rugged adaptive landscapes, where lineages can get stuck on local fitness peaks, unable to cross the valleys towards higher ones. The emergence of Alpha might have originated in the context of long-term chronic infection, through a burst of mutations that allowed it to reach a higher fitness peak, that may have been difficult to attain through sequential accumulation of mutations in immunocompetent hosts<sup>34</sup>.

The epidemiological success of some lineages can also be assessed more directly through growth rate estimates<sup>28</sup>, aided by independent observations in different epidemiological settings, notwithstanding the challenging of adjusting for highly unequal sampling. Growth rate estimates must be carefully tested for confounders as initial increases due to an adaptive advantage may be difficult to disentangle from neutral processes which can also give rise to marked lineage dynamics (**Figure 2**). Finally, selection screens based on patterns of non-synonymous and synonymous mutations<sup>51</sup> and homoplasmy-based metrics<sup>25</sup> may hold scope to predict the adaptive potential of a variant based on its constituent mutations.

#### *Ongoing evolution of SARS-CoV-2*

The limited evidence for early adaptation detected in SARS-CoV-2 suggests some degree of pre-existing adaptation to human infection, either through existing generalist mechanisms of host evasion or because the earliest lineages of SARS-CoV-2 had already acquired some key mutations favouring transmissibility in humans by the time the first sequences were generated. Nevertheless, the emergence of VoCs with clearly higher transmissibility (Alpha, Delta) are obvious signatures of adaptation of SARS-CoV-2 to its human host. The exact mechanisms underlying increased transmissibility still remain poorly quantified and many of the characteristic mutations deemed as concerning (**Figure 3**) are also observed in other lineage backgrounds, often well predating the emergence of VoCs, and without apparent fitness effects. Besides epistatic interactions between mutations, a further complicating factor stems from the increasing rates of immunisation due to vaccination and natural infection which alters the immunological environment of the virus as the pandemic progresses, thereby possibly modifying the fitness effect of immune-escape mutations.

Virulence, defined as the disease-induced mortality rate for infected hosts, is even more challenging to predict. While it is sometimes assumed that viral pathogens systematically become attenuated following a host jump, this is only necessarily true for vertically (i.e. mother to child) inherited pathogens. In the latter situation the fitness of the pathogens and the host are intimately correlated so that high virulence would entail a fitness cost. Under horizontal (epidemic) transmission, a pathogen can become more or less lethal depending on the correlation between transmissibility and virulence. For example, higher viral replication rates can be favoured as long as they allow a lineage to spread more easily, even if this entails a cost to the host. Given the moderate host mortality and ~50% of transmission being pre-

symptomatic, the selective pressure on SARS-CoV-2 to evolve towards intrinsic lower virulence is expected to be weak. This being said, morbidity and mortality are predicted to eventually plummet as an increasing fraction of the population acquires some immunity through vaccination and prior infection.

Likely the most significant aspect to pandemic management is the degree of evolution of SARS-CoV-2 towards altered antigenicity, the timescales and consistency of which can be difficult to predict even for richly studied viruses such as seasonal influenza. Concerns surrounding waning immunity and escape variants have in part been driven by studies of reinfection by other human seasonal coronaviruses. Two major hypotheses could explain frequent reinfections: rapidly waning host immunity; or pathogen antigenic evolution leading to immune escape. Neutralisation assays (**Box 1**) provide support for the latter hypothesis in alphacoronavirus 229E. Using serum collected between 1985 and 1990, and a series of 229E spike sequences sampled at eight-year intervals, Eguia and colleagues showed that antigenic evolution over the course of a decade led to almost complete escape from antibody neutralisation after 8-17 years. This suggests that vaccines to SARS-CoV-2 may need to be reformulated periodically, as circulating SARS-CoV-2 lineages evolve in response to host immunity<sup>77</sup>. Vaccine escape, however, is rarely a binary phenotype and a gradual loss of antibody recognition through the evolution of new viral strains is to be expected rather than an instantaneous loss of efficacy. Moreover, cellular immunity is still predicted to provide some protection against most severe symptoms in many re-infections. Thus, while updated vaccinations will be required at least for the most at-risk, this should be a largely manageable challenge aided by genomic surveillance and the capability to readily update modern vaccines.

## **Conclusions**

Genomic tracking of SARS-CoV-2 has been instrumental in monitoring the course of the COVID-19 pandemic and will become a vital component supporting ongoing public health, in particular informing on the vaccine update schedule. There is little doubt that SARS-CoV-2 will become an endemic circulating pathogen which will continue to undergo adaptation to human cellular and immune defences. However, early identification of new variants, coupled with a richer understanding of the mutational forces underlying changing patterns of transmissibility, virulence and antigenicity could reduce global morbidity and mortality to a fraction of that experienced during the pandemic phase of COVID-19. Additionally, the tremendous progress and lessons learnt from fighting SARS-CoV-2 hold promise to reduce future threats to global health and support preparedness for future epidemics.

## **Acknowledgments**

We gratefully acknowledge all providers and originators contributing to GISAID, on which **Figure 1** and **Figure 2** is based. An acknowledgements table is provided at <https://figshare.com/s/77acedc784d79babb957>. We particularly acknowledge the COG-UK consortium who have publicly released a large number of SARS-CoV-2 genomes from the UK. In addition, we thank the developers and maintainers of the outbreak.info and covariants.org platforms. Funding: L.vD is supported by a UCL Excellence Fellowship. D.R is supported by an NIHR Precision AMR award. C.J.H is supported by Wellcome collaborative grant 204870/Z/16/Z. F.B acknowledges support from a BBSRC (equipment grant BB/R01356X/1).

**Declaration of interest:** none.

## References and Recommended Reading

\*special interest

\*\*outstanding interest

1. Fraser, C. *et al.* Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science (80-. )*. **324**, 1557–1561 (2009).
2. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
3. Faria, N. R. *et al.* Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* **8**, 97 (2016).
4. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
5. Holmes, E. C. *Novel 2019 coronavirus genome - SARS-CoV-2 coronavirus - Virological.* <https://virological.org/t/novel-2019-coronavirus-genome/319> (2020).  
**\*\*Virological post describing the release of the first SARS-CoV-2 genome.**
6. van Dorp, L. *et al.* Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.* **83**, 104351 (2020).  
**\*Initial characterisation of the distribution of recurrent mutations in the early SARS-CoV-2 population.**
7. Pekar, J., Worobey, M., Moshiri, N., Scheffler, K. & Wertheim, J. O. Timing the SARS-CoV-2 index case in Hubei province. *Science (80-. )*. eabf8003 (2021)  
doi:10.1126/science.abf8003.
8. Kumar, S. *et al.* An evolutionary portrait of the progenitor SARS-CoV-2 and its dominant offshoots in COVID-19 pandemic. *Mol. Biol. Evol.* (2021)  
doi:10.1093/molbev/msab118.
9. La Rosa, G. *et al.* SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. *Sci. Total Environ.* **750**, 141711 (2021).
10. du Plessis, L. *et al.* Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science (80-. )*. **10**, eabf2946 (2021).  
**\*Characterisation of the SARS-COV-2 transmission clusters introduced into the UK in the first pandemic wave.**
11. Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and North America. *Science* (2020) doi:10.1126/science.abc8169.
12. Qutob, N. *et al.* The genomic epidemiology of SARS-CoV-2 in Palestine. *bioRxiv* 2020.10.26.355677 (2020) doi:10.1101/2020.10.26.355677.
13. Oude Munnink, B. B. *et al.* Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* **26**, 1405–1410 (2020).
14. da Silva Filipe, A. *et al.* Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland. *Nat. Microbiol.* **6**, 112–122 (2021).
15. Hodcroft, E. B. *et al.* Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv* vol. 12 16 (2020).
16. Aggarwal, D. An integrated analysis of contact tracing and genomics to assess the efficacy of travel restrictions on SARS-CoV-2 introduction and transmission in England from June to September, 2020. *medRxiv* 2021.03.15.21253590 (2021)  
doi:10.1101/2021.03.15.21253590.
17. Meredith, L. W. *et al.* Rapid implementation of SARS-CoV-2 sequencing to

- investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* **20**, 1263–1272 (2020).
18. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob. Challenges* **1**, 33–46 (2017).
  19. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, 30494 (2017).
  20. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
  21. Poon, A. F. Y. CoVizu: Rapid analysis and visualization of the global diversity of SARS-CoV-2 genomes - SARS-CoV-2 coronavirus / Software and Tools - Virological. *Virological* <https://virological.org/t/covizu-rapid-analysis-and-visualization-of-the-global-diversity-of-sars-cov-2-genomes/678> (2021).
  22. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 1–5 (2020) doi:10.1038/s41564-020-0770-5.  
**\*\*Publication describing the dynamic PANGO lineage nomenclature tool.**
  23. Hodcroft, E. B. *et al.* Want to track pandemic variants faster? Fix the bioinformatics bottleneck. *Nat.* 2021 5917848 **591**, 30–33 (2021).
  24. van Dorp, L. *et al.* No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat. Commun.* **11**, 5986 (2020).
  25. Richard, D. *et al.* A phylogeny-based metric for estimating changes in transmissibility from recurrent mutations in SARS-CoV-2. *bioRxiv* 2021.05.06.442903 (2021) doi:10.1101/2021.05.06.442903.
  26. Yatish Turakhia, *et al.* Ultrafast Sample Placement on Existing Trees (USHER) Empowers Real-Time Phylogenetics for the SARS-CoV-2 Pandemic. *Nat. Genet.* **1348**, (2021).
  27. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e19 (2020).
  28. Volz, E. *et al.* Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* **184**, 64-75.e11 (2021).
  29. Thomson, E. C. *et al.* Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**, 1171-1187.e20 (2021).
  30. Munnink, B. B. O. *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science (80-. ).* (2020) doi:10.1126/SCIENCE.ABE5901.  
**\*Study documenting zoonotic and anthroponotic transmissions of SARS-CoV-2 to farmed minks.**
  31. van Dorp, L. *et al.* Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. *bioRxiv* 2020.11.16.384743 (2020) doi:10.1101/2020.11.16.384743.
  32. Hoffmann, M. *et al.* SARS-CoV-2 mutations acquired in mink reduce antibody-mediated 1 neutralization 2 3. *bioRxiv* 2021.02.12.430998 (2021) doi:10.1101/2021.02.12.430998.
  33. Garcia-Beltran, W. F. *et al.* Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372-2383.e9 (2021).
  34. Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (2020).  
**\*\*Virological post providing initial description of the B.1.1.7 VoC.**

35. Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* 2020.12.21.20248640 (2020)  
doi:10.1101/2020.12.21.20248640.  
**\*\*Initial publication describing the B.1.351 VoC first identified in South Africa.**
36. Faria, N. R. *et al.* Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. eabh2644 (2021) doi:10.1126/science.abh2644.  
**\*\*Initial publication describing the P.1 VoC first linked to infections in Manaus, Brazil.**
37. Nelson, G. *et al.* Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escap. *bioRxiv* 2021.01.13.426558 (2021)  
doi:10.1101/2021.01.13.426558.
38. Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e20 (2020).
39. Zahradník, J. *et al.* SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv* 2021.01.06.425392 (2021)  
doi:10.1101/2021.01.06.425392.
40. Volz, E. *et al.* Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv* 2020.12.30.20249034 (2021)  
doi:10.1101/2020.12.30.20249034.
41. Cele, S. *et al.* Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* **593**, 142–146 (2021).
42. Hoffmann, M. *et al.* SARS-CoV-2 variants B.1.351 and B.1.1.248: Escape from therapeutic antibodies and antibodies induced by infection and vaccination. *Cell* **184**, 2384–2393 (2021).
43. Wibmer, C. K. *et al.* SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* 1–4 (2021) doi:10.1038/s41591-021-01285-x.
44. Wu, K. *et al.* mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. *bioRxiv Prepr. Serv. Biol.* 2021.01.25.427948 (2021) doi:10.1101/2021.01.25.427948.
45. Kymie, C. *et al.* Genomic evidence of a SARS-CoV-2 reinfection case with E484K spike mutation in Brazil. <https://www.gisaid.org/> (2021).
46. Bager, P. *et al.* Increased Risk of Hospitalisation Associated with Infection with SARS-CoV-2 Lineage B.1.1.7 in Denmark. *SSRN Electron. J.* (2021)  
doi:10.2139/ssrn.3792894.
47. Davies, N. G. *et al.* Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 1–5 (2021) doi:10.1038/s41586-021-03426-1.
48. Santos De Oliveira, M. H., Lippi, G. & Henry, B. M. Sudden rise in COVID-19 case fatality among young and middle-aged adults in the south of Brazil after identification of the novel B.1.1.28.1 (P.1) SARS-CoV-2 strain: analysis of data from the state of Parana. *medRxiv* 2021.03.24.21254046 (2021) doi:10.1101/2021.03.24.21254046.
49. Zhang, W. *et al.* Emergence of a novel SARS-CoV-2 strain in Southern California, USA. *medRxiv* 2021.01.18.21249786 (2021) doi:10.1101/2021.01.18.21249786.
50. Mlcochova, P. *et al.* SARS-CoV-2 B.1.617.2 Delta variant emergence and vaccine breakthrough 1 2. *bioRxiv* 2021.05.08.443253 (2021) doi:10.1101/2021.05.08.443253.
51. Martin, D. P. *et al.* The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2 selective landscape.

- medRxiv* (2021) doi:10.1101/2021.02.23.21252268.
52. Teruel, N., Mailhot, O. & Najmanovich, R. J. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. *bioRxiv* 2020.12.16.423118 (2020) doi:10.1101/2020.12.16.423118.
  53. Greaney, A. J. *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463-476.e6 (2021).
  54. Andreano, E. *et al.* SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv* 2020.12.28.424451 (2020) doi:10.1101/2020.12.28.424451.
  55. Wang, Z. *et al.* mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616 (2021).
  56. Li, Q. *et al.* The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. (2020) doi:10.1016/j.cell.2020.07.012.
  57. Motozono, C. *et al.* An emerging SARS-CoV-2 mutant evading cellular immunity and increasing viral infectivity. doi:10.1101/2021.04.02.438288.
  58. Starr, T. N., Greaney, A. J., Dingens, A. S. & Bloom, J. D. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Reports Med.* **2**, 100255 (2021).
  59. Brown, J. C. *et al.* Increased transmission of SARS-CoV-2 lineage B.1.1.7 (VOC 202012/01) is not accounted for by a replicative advantage in primary airway cells or antibody escape. *bioRxiv* 2021.02.24.432576 (2021) doi:10.1101/2021.02.24.432576.
  60. McCarthy, K. R. *et al.* Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science (80-. )*. **371**, 1139–1142 (2021).  
**\*Characterisation of notable deletions in SARS-CoV-2 and their role in antibody escape.**
  61. Kemp, S. A. *et al.* SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592**, 277–282 (2021).
  62. Mccallum, M. *et al.* N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *bioRxiv* 2021.01.14.426475 (2021) doi:10.1101/2021.01.14.426475.
  63. Kemp, S. *et al.* Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion  $\Delta$ H69/V70. *bioRxiv* 2020.12.14.422555 (2020) doi:10.1101/2020.12.14.422555.
  64. Meng, B. *et al.* Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep.* **35**, 109292 (2021).
  65. Xia, H. *et al.* Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep.* **33**, 108234 (2020).
  66. Matthew Parker, A. D. *et al.* Altered Subgenomic RNA Expression in SARS-CoV-2 B.1.1.7. *bioRxiv* 2021.03.02.433156 (2021) doi:10.1101/2021.03.02.433156.
  67. Young, B. E. *et al.* Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet (London, England)* **396**, 603–611 (2020).
  68. Chandrashekar, A. *et al.* SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science (80-. )*. **369**, 812–817 (2020).
  69. Lumley, S. F. *et al.* Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers. *N. Engl. J. Med.* **384**, 533–540 (2021).
  70. Staub, T. *et al.* Case series of four re-infections with a SARS-CoV-2 B.1.351 variant, Luxembourg, February 2021. *Eurosurveillance* **26**, 2100423 (2021).
  71. Schmidt, F. *et al.* Measuring SARS-CoV-2 neutralizing antibody activity using

- pseudotyped and chimeric viruses. *J. Exp. Med.* **217**, (2020).
72. Ferreira, I. *et al.* SARS-CoV-2 B.1.617 emergence and sensitivity to vaccine-elicited antibodies. *bioRxiv* 2021.05.08.443253 (2021) doi:10.1101/2021.05.08.443253.
  73. Collier, D. A. *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136 (2021).
  74. Gonzalez, C. *et al.* Live virus neutralisation testing in convalescent patients and subjects vaccinated against 19A, 20B, 20I/501Y.V1 and 20H/501Y.V2 isolates of SARS-CoV-2. *medRxiv* 2021.05.11.21256578 (2021) doi:10.1101/2021.05.11.21256578.
  75. O'Reilly, K. M., Allen, D. J., Fine, P. & Asghar, H. The challenges of informative wastewater sampling for SARS-CoV-2 must be met: lessons from polio eradication. *The Lancet Microbe* vol. 1 e189–e190 (2020).
  76. Fujino, T. *et al.* Novel SARS-CoV-2 variant in travelers from Brazil to Japan. *Emerg. Infect. Dis.* **27**, 1243–1245 (2021).
  77. Eguia, R. *et al.* A human coronavirus evolves antigenically to escape antibody immunity. *PLOS Pathog.* **17**, e1009453 (2021).
- \*\*Study of antigenic evolution of seasonal endemic coronavirus spike over 8 years suggests timeframe to antibody escape.**