Journal of the Pediatric Infectious Diseases Society

SUPPLEMENT ARTICLE

Jumping a Moving Train: SARS-CoV-2 Evolution in Real Time

Ahmed M. Moustafa[1,2] **and Paul J. Planet**[1,3,4]

[1]Division of Pediatric Infectious Diseases, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA, [2]Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA, [3]Department of Pediatrics, Perelman College of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA, and [4]Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York, USA

The field of molecular epidemiology responded to the SARS-CoV-2 pandemic with an unrivaled amount of whole viral genome sequencing. By the time this sentence is published we will have well surpassed 1.5 million whole genomes, more than 4 times the number of all microbial whole genomes deposited in GenBank and 35 times the total number of viral genomes. This extraordinary dataset that accrued in near real time has also given us an opportunity to chart the global and local evolution of a virus as it moves through the world population. The data itself presents challenges that have never been dealt with in molecular epidemiology, and tracking a virus that is changing so rapidly means that we are often running to catch up. Here we review what is known about the evolution of the virus, and the critical impact that whole genomes have had on our ability to trace back and track forward the spread of lineages of SARS-CoV-2. We then review what whole genomes have told us about basic biological properties of the virus such as transmissibility, virulence, and immune escape with a special emphasis on pediatric disease. We couch this discussion within the framework of systematic biology and phylogenetics, disciplines that have proven their worth again and again for identifying and deciphering the spread of epidemics, though they were largely developed in areas far removed from infectious disease and medicine.

Key words. COVID-19; lineages; nomenclature; VOC.

The SARS-CoV-2 virus was first reported as a cluster of pneumonia cases in Wuhan, China on December 31, 2019 with the first symptomatic case presenting on December 1 [1], although there have been unconfirmed reports of cases from November [2]. The first case outside China was reported in a traveler from Wuhan to Thailand [3], and by the end of January, cases had been reported in 19 countries outside of China [4]. While the initial cases outside of China were mostly linked to travel, sustained community spread outside of China is thought to have begun globally in February [5, 6].

The virus itself was identified in early January as a novel betacoronavirus (sarbecovirus) that was sequenced directly from respiratory specimens [7, 8], and the first SARS-CoV-2 whole genome was submitted to GenBank on January 5 (released January 12). Subsequent sequencing of the virus using targeted PCR amplification confirmed the association of this virus with the initial cases [8]. The whole genome was critical in allowing the early development of PCR-based testing [9] and genomic amplification (https://artic.network/), as well as beginning to understand the biology of the virus even before the world scientific community had physical access to the virus for study [10]. For instance, the similarity of the receptor-binding domain (RBD) to the SARS-CoV-1 virus suggested that it bound to the ACE2 receptor on human cells [7, 11], which was soon after confirmed [8, 12]. It also suggested a close phylogenetic relationship with circulating bat coronaviruses [7, 8].

While early phylogenetic analysis tied the viral sequence strongly to other known viruses from horseshoe bats (Genus: *Rhinolophus*) [7, 8], the position in the phylogenetic tree differed based on which genes were used to build the phylogeny [7]. In particular, the spike protein gene (S) was highly similar to and grouped with the S-gene from SARS-CoV-1 while other genes suggested other closest relatives such as the bat virus RaTG13 [7]. Other sequence traits, especially in the variable loop region of the RBD of the spike protein, are similar to coronaviruses isolated from pangolins, leading to the hypothesis that there had been recombination between viruses between different hosts [7, 13] and that pangolins might be an intermediate or proximate host to the pandemic virus [11, 14–18]. Recent structure and binding studies of bat, pangolin, and SARS-CoV-2 spike proteins showed high affinity for pangolin and human ACE2 proteins and low affinity for bat ACE2 [19–21], but it may be that the protein differences expand binding to multiple different hosts [19]. Another study challenged the recombination and pangolin host hypothesis, arguing that the RaTG13 may have acquired its variable loop region from another sarbecovirus, making the RBD possessed by SARS-CoV-2 the ancestral form [22]. At this time, the host

Corresponding Author: Paul J. Planet, MD, PhD, Division of Infectious Disease, The Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Suite 1202, Philadelphia, PA 19104, USA
E-mail: planetp@chop.edu.

Journal of the Pediatric Infectious Diseases Society 2021;XX(XX):1–10
© The Author(s) 2021. Published by Oxford University Press on behalf of The Journal of the Pediatric Infectious Diseases Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
DOI: 10.1093/jpids/piab051

source of SARS-CoV-2 is still unclear, however, if the virus has been circulating in bat populations for decades as has been predicted by some analyses [22], comprehensive sampling of bat populations may reveal important observations.

Another novel characteristic noted from the whole-genome sequences [11, 23] was a polybasic cleavage site at the junction of the 2 major spike protein domains, a furin cleavage site, in the spike protein that was shown to have a role in SARS-CoV-2 replication, transmission and pathogenesis [24, 25], although some evidence suggests that increased spike proteolysis is associated with reduced viral entry and infectivity [26–28]. However, cleavage of the spike protein seems to be an adaptation that enhances viral entry, cell-cell fusion, and host range in many viruses such as Influenza, MERS, and SARS-CoV-1 [29–33]. Another amino acid feature flanking the furin cleavage site is the prediction of sites for O-linked glycan modification that has recently been shown to impact viral fusion and entry [28] and has been proposed to provide an immunological shield at this site [11, 34, 35].

There has been speculation that the suite of traits that likely enhance the virulence, transmissibility, and host range of this virus somehow demonstrates that the virus was genetically engineered in a laboratory. There are many arguments against this including convincing and careful analyses about the relationship with viruses from bats [8, 13, 22, 36] and, that despite these close relationships, the virus sequence is still substantially different, with variable positions distributed throughout its genome, from any virus known in the literature or used in a laboratory. Overall, it seems extremely unlikely that the virus is derived from a laboratory [11, 22, 35, 37, 38].

MODES OF EVOLUTION AND SELECTION

SARS-CoV-2 is an RNA virus and therefore has a relatively fast mutation rate. Most estimates of global mutation rates range from 5.44E-4 to 1.22E-3, which is approximately 1-2 nucleotide changes per month in any given lineage [22, 39–42]. The genome also undergoes insertions and deletions (indels) at a less well-determined rate [43].

This rate, however, is most likely to be the observed rate overall in viruses that transmit. In actuality, mutational change is likely more frequent, and in any given infection not all viruses will be exactly the same. This variation within each infection is something that has been studied more deeply in other viruses such as HepCV, HepBV, and HIV to name a few, where the diversity of mutations in the infecting virus has been shown to have a profound impact on viral fitness and evolution [44].

Here, the concept of "quasispecies" is important. A quasispecies is a population of organisms that are all very closely related to one another, usually differing by a small number of mutations, that have a collective fitness and can be acted on by selection as a unit [44]. That is, it does not matter how fit the most fit virus is in the population, but it does matter how fit the population is overall. RNA viruses are notorious for behaving like quasispecies because they have high mutation rates and short genomes [44], and both SARS-CoV-1 [45] and MERS [46] have been documented to exist in quasispecies. Therefore it seems likely that quasispecies behavior may be important for SARS-CoV-2. The number of studies documenting quasispecies behavior has been increasing [47–56], but the impact of quasispecies on virulence, immune evasion, and spread is not well understood. Current sequencing methods cannot easily detect quasispecies since they rely on direct amplification of the viral genome from clinical specimens and then consensus sequence building. The genome is, therefore, representative of the dominant forms of the virus in any given infection, and only very deep sequencing can reveal nucleotides that may be there as minor variants. Future studies will need to sequence very deeply and/or use single genome capture [57, 58] or long read strategies [53] to understand quasispecies for SARS-CoV-2.

The role of recombination (where distantly related viruses swap portions of their genomes) is also not clear, though it has been proposed as important for the origins of the virus [36, 59–62] and has been shown in some circulating lineages [63, 64]. Indeed, a naming convention has been proposed for them within the Pango system, where the highest level lineage is preceded by an X [65]. Recombination is worrisome because it provides another way for the virus to gain beneficial mutations without having to rely on the random process of mutation. Recombination, of course, relies on 2 unrelated viruses infecting the same cell at the same time. Evidence for co-infections of distinct viruses have been limited so far [63, 66, 67], but it is not clear (as mentioned above) that the most common, current sequencing techniques could detect infections with different viruses.

One important aspect of SARS-CoV-2 evolution that we have witnessed in real time is the recurrent generation of the same mutation in different genetic backgrounds [68–71]. This could arise from recombination or by the same random event happening more than once in different genomes. When this type of event happens, and we observe different viruses with the same mutation spreading well in community, it is one of the best kinds of evidence that a particular mutation has a positive impact on viral fitness. True evolutionary convergences or parallelisms (or "homoplasies" in systematic biology terminology) are therefore goldmines for understanding viral biology and identifying variants of concern.

However, detecting truly homoplastic changes can be difficult. Some mutations may look like they have occurred in different backgrounds just because they have stayed the same as the other parts of the viral genome changed. Indeed, this has been proposed as the evolutionary pattern that resulted in the diversity in the variable loop region [22]. In systematic biology, this would be called a "symplesiomorphy," a mutation that viruses share because their ancestors had it as well. Symplesiomorphies provide much weaker evidence for variant fitness, although their

conservation over time may signal some benefit. The best way to distinguish between convergence and symplesiomorphy is to understand how viruses are related to each other on a phylogenetic tree, therefore phylogeny is not only important for taxonomy but can be critical for identifying functional mutations.

LINEAGES, VARIANTS, AND NOMENCLATURE

As viral lineages have spread around the globe they have changed enough that they can be distinguished from one another and traced geographically. In this publication, we use the term "variant" to refer to viral genomes that can be distinguished based on one or more specific genomic changes. It is worthwhile noting that variants belonging to the same group may not be closest relatives. We use the term "lineage" to refer to a group of viruses thought to be related to each other by descent from a common ancestor. A "clade" is the group of organisms that includes *all* of the descendants of a common ancestor, whereas a lineage may only include some descendants. It is also important to note that variants, lineages, and clades are designated here only by their genotype without implying any functional or phenotypic changes.

Nomenclature Systems and Classification Tools

The rapidly evolving genomic forms of the virus required quick elaboration of nomenclature systems, and a strong case was made from the beginning that these systems should be phylogenetically based [72]. Borrowing from previous propositions to tie nomenclature closely to phylogeny (eg, http://phylonames.org/code/), Rambaut and colleagues [72] put forward the use of a string of numbers and letters separated by periods, in which each lineage contains a nested list of its ancestral lineages in its name. Thus, B.1.1.7 suggests that this lineage is the seventh lineage derived from a lineage B.1.1, which itself is the first lineage derived from a lineage called B.1. This Pango nomenclature system, which is implemented in the program (Pangolin) (https://github.com/cov-lineages/pangolin) is intuitive and informative, such that we know that B.1.1.7 is more closely related to B.1.1.6 than it is to B.2.1.7 without having to look it up on a phylogenetic tree. One downside to this kind of nomenclature is that each new lineage must be defined by manual curation of the sequences, requiring a central authority that controls and certifies new lineages. While the Pango system has been widely adopted, some other nomenclature systems have also been used such as Nextstrain and GISAID [73, 74], all of which require manual curation to define which clades on the phylogenetic tree should be named. We recently proposed an automated system of nomenclature based on whole-genome multilocus sequence typing (wgMLST) that defines groups based on clear-cut properties of genomes on a minimum spanning phylogenetic tree. This technique, called GNUVID [75], provides an adjunctive technique that can help add granularity to the Pango system.

Because it is costly to re-calculate the entire phylogenetic tree for every new sequence that is obtained several techniques have been developed to assign new sequences to lineages based on how closely they match other sequences in the database. Early techniques used alignment and measurements of similarity to viruses already included in a specific group or lineage [73, 76, 77]. More recent techniques have used machine learning, which can also provide estimates of confidence in the taxonomic assignment [72, 75].

Phylogeny and Phylodynamics

As a natural way to explain variation, heredity, relatedness, and change over time, phylogenies form the backbone of evolutionary biology. From a practical standpoint, phylogenies not only offer the ability to classify genomes into lineages but can also identify transmission and importation events and reconstruct general patterns of spread. Phylogenetic network and tree-based approaches have been used extensively in tracing the emergence of SARSA-CoV-2. Phylogenies can also be used to infer rates of evolution and identify specific sequence changes associated with biological properties of the virus (eg, increased transmissibility or virulence). Phylodynamic approaches [78] add population genetic models and epidemiologic data to the phylogenetic framework and can be used to estimate parameters such as rates of spread, transmissibility, and effective population size. Geospatial data can also be incorporated to estimate the spread of lineages in space.

Without active surveillance and phylodynamic modeling of viral lineages, it would have been very difficult, if not impossible, to identify certain viral lineages as more transmissible. These types of analyses strongly suggested that the B.1.1.7 lineage was 50% more transmissible than other circulating clones [79], and helped inform policy decisions around containment of that lineage, unfortunately too late to stop the global spread of that lineage.

As practiced by evolutionary biologists, phylogenetics has mostly been built on events that happened in the distant past. A problematic by-product of building phylogenies and taxonomical schemes in real time, as a virus evolves, is that it is almost inescapable that some groupings will not represent all the descendants of a single ancestral virus. In other words, as new groups arise from older groups, the older groups will not necessarily remain "monophyletic." Taxonomic monophyly is strongly preferred because it is not arbitrary, whereas a paraphyly (a group that has a common ancestor and only some of its descendants) depends upon drawing an arbitrary line in the sand, including only some descendants and excluding others. For instance, the moment that B.1.1 was named and recognized as its own clade, the B.1 clade became paraphyletic. This means that there are likely some members of B.1 that are more closely related to B.1.1 than others. Because of this B.1.1 and B.1 are not really comparable units, a taxonomic

issue that needs to be kept in mind when comparing named lineages.

TRACING THE ORIGINS OF LINEAGES AND TRACKING THE SPREAD OF THE VIRUS

Global Patterns

The pandemic might be divided into a few critical mutational events that have defined the spread of the virus around the world. After its initial emergence in Wuhan, the first event was a mutation at the D614G position in the spike protein that led to a lineage of viruses with higher transmission rates [27]. The biological basis of increased transmission seems to be an increased affinity and cell entry [80–83] and potentially increased surface spike protein density such that the virus has more possible ligands for targeting the ACE2 receptor expressed on host cells [84]. Given the available data, this mutation likely arose in Asia (it was first sequenced in China) and then spread to Europe and around the world. While the very earliest global cases of COVID-19 did not have the D614G variant, viruses with this mutation very soon came to dominate the early phase of the pandemic in many countries. The earliest outbreaks in Italy, Spain and the United Kingdom all were dominated by this mutation [27]. In addition, the lineage that dominated the early New York spike in March and April of 2020 also had D614G [85].

The next major phase started with the introduction of the B.1.1.7, which was recognized by an unusual suite of 17 changes (Table 1). This lineage was remarkable for 2 major reasons. First, it seemed to be spreading much faster than other viruses, a conclusion that could only be reached because of a concerted effort to sequence viruses in the United Kingdom [79, 86] and phylodynamic integration with community testing data that showed PCR S-gene target failures, or "S-drop," even when other parts of the viral genome were detected. S-drop is due to a deletion mutation in the S-gene (del69-70) [86]. Regions of the United Kingdom where there were faster increases in cases were identified to have higher rates of B.1.1.7, and S-drop cases seemed to increase faster than non-S-drop cases in the same locations in England [86]. In addition, measurements of infection of contacts by index cases were higher for B.1.1.7 compared to other circulating strains [87].

The second unusual feature of B.1.1.7 was a high number of putatively adaptive mutations, leading to the hypothesis that the origin of B.1.1.7 may have been in a prolonged infection in a single host [89–93]. The del69-70 mutation in the spike protein has been implicated in increased spread [94, 95]. Likewise, a N501Y mutation in the spike protein has also been shown by modeling and in vitro to more avidly bind the ACE2 receptor [70, 96–98]. The mutation, P681H that is immediately adjacent to the furin cleavage site has also been suggested to increase transmission rates [25, 26].

The origin of the B.1.1.7 lineage is unclear, but the earliest B.1.1.7 genome was reported from the United Kingdom on September 2020 and was the dominant strain in the United Kingdom by December 2020 [86]. It very quickly went on to become dominant in several other European countries and in Israel [99]. More recently starting in March 2021, B.1.1.7 has become the predominant lineage in the United States, with an extremely rapid increase across the country, coinciding with a major vaccination effort.

The N501Y mutation is likely convergent across several lineages, and 2 other important lineages with this mutation, B.1.351 and P.1, also emerged around the same time as B.1.1.7 in South Africa and Brazil respectively. These 2 lineages also had a worrisome mutation, E484K, that has been shown to enhance the escape of neutralizing antibodies in vitro [100–102] and may be linked to lower efficacy for vaccines [103–106]. Interestingly, the B.1.351 lineage appears to have greater reductions in neutralizing antibodies than the P.1 lineage [107]. This same mutation has also surfaced in several B.1.1.7 isolates from around the world [108] though only a small number of studies have evaluated the impact on neutralizing antibody escape for this variant.

Currently, the B.1.617 lineage and its sublineages (B.1.617.1, B.1.617.2, and B.1.617.3) are circulating in a massive spike of cases in India, and this lineage is already spreading globally. The critical mutations in this lineage appear to be L452R and E484Q, although B.1.617.2 lacks the latter and instead has T478K. E484Q, like E484K, is thought to reduce antibody efficacy and neutralizing antibodies. L452R variants in the B.1.427/429 lineage have recently been reported to be more transmissible and infective as well as less susceptible to neutralizing antibodies [109]. Less is known about the T478K but it has been noted to be increasing dramatically since January 2021 in Mexico and North America [110]. Interestingly, B.1.617.2 appeared to be expanded rapidly in the United Kingdom in April and May [111].

Local Patterns

One important aspect of whole-genome sequencing and classification has been the ability to detect new variants as they arise or are imported to a new place. Early in the pandemic, whole genomes allowed researchers to estimate the numbers and sources of introductions of the virus [112–114]. Since these initial observations, a striking realization has been the huge number of predicted importations and exportations between countries and across continents, even with significant restrictions on travel. A close look at any part of the SARS-CoV-2 phylogeny shows multiple exportations and importations. Reliable quantification of the flux across borders is highly dependent on the density and breadth of sampling. A detailed analysis was possible only in the UK, where there has been a strong commitments to sampling and sequencing genomes [113]. In this study, greater than 1000 viral introductions were detected prior to the lockdown in March 2020 and then dropped significantly, but were not completely controlled with the lockdown [113]. The

Table 1. Characteristics of the SARS-CoV-2 variants of interest and concern.

PANGO Lineage	B.1.525	B.1.526	B.1.526.1	B.1.617	B.1.617.1	B.1.617.3	P.2	B.1.1.7	B.1.351	B.1.427	B.1.429	B.1.617.2	P.1
Classification	VOI	VOI	VOI	VOI	VOI	VOI	VOI	VOC	VOC	VOC	VOC	VOC	VOC
WHO label	Eta	Iota	-	-	Kappa	-	Zeta	Alpha	Beta	Epsilon	Epsilon	Delta	Gamma
Nextstrain name	20A/S:484K	20C/S:484K	20C	20A	20A/S:154K	20A	20J	20I/501Y.V1	20H/501.V2	20C/S:452R	20C/S:452R	20A/S:478K	20J/501Y.V3
First detected	UK/Nigeria	NY	NY	India	India	India	Brazil	UK	South Africa	CA	CA	India	Japan / Brazil
No. of spike mutations	8	3–7	6–8	3	7–8	7	3–4	10–13	10	4	4	9–10	11
RBD mutations	E484K	(S477N*) (E484K*)	L452R	L452R E484Q	L452R E484Q	L452R E484Q	E484K	N501Y	K417N E484K N501Y	L452R	L452R	L452R T478K	K417T E484K N501Y
Other spike mutations	A67V, 69del, 70del, 144del, D614G, Q677H, F888L	(L5F*), T95I, D253G, D614G, (A701V*)	D80G, 144del, F157S, D614G, (T791I*), (T859N*), D950H	D614G	(T95I*), G142D, E154K, D614G, P681R, Q1071H	T19R, G142D, D614G, P681R, D950N	(F565L*), D614G, V1176F	69del, 70del, 144del, (E484K*), (S494P*), A570D, D614G, P681H, T716I, S982A, D1118H (K1191N*)	D80A, D215G, 241del, 242del, 243del, D614G, A701V	S13I, W152C, D614G	S13I, W152C, D614G	T19R, (G142D*), 156del, 157del, R158G, D614G, P681R, D950N	L18F, T20N, P26S, D138Y, R190S, D614G, H655Y, T1027I
Transmission	-	-	-	-	-	-	-	50% increased	50% increased	20% increased	20% increased	Increased	-
Neutralization by convalescent and vaccine sera	Potential reduced	Reduced	Potential reduced	Slightly reduced[a]	Potential reduced[a]	Potential reduced[a]	Reduced[a]	Minimal impact	Reduced	Reduced	Reduced	Potential reduced[a]	Reduced
Neutralization by some EUA monoclonal antibody treatments	Potential reduced	Reduced[b]	Potential reduced	Potential reduced	Potential reduced	Potential reduced	Potential reduced	No Impact	Significant decrease[b]	Modest decrease[b]	Modest decrease[b]	Potential reduced	Significant decrease[b]

Abbreviations: WHO, World Health Organization; EUA, Emergency Use Authorization; CA, California; NY, New York; RBD, receptor-binding domain; UK, United Kingdom; VOC, variant of concern; VOI, variant of interest.
This table is reproduced from data available on CDC website [88].
(*) means found in some sequences but not all.
[a]The reported reductions in B.1.617, B.1.617.1, B.1.617.2, B.1.617.3, and P.2 are for neutralization by vaccine sera. No data for convalescent sera.
[b]This reduction in susceptibility to the bamlanivimab and etesevimab monoclonal antibody combination treatment, however, other monoclonal antibody treatments are available.

number of introductions in the early pandemic into countries with much less genomic sampling, like the United States, will likely never be known, but new sampling efforts may increase our knowledge of importations and allow for rational approaches to curbing spread.

The diversity of viruses in any given location is determined by the rate at which new viruses arise by mutation (or possibly recombination), the number of introductions, and the extinction of lineages if they fail to transmit. Each of these factors depends on the ability of the virus to transmit and persist in the community, and therefore viral diversity is reflective of overall viral fitness. Measurements of circulating diversity might be useful tools for monitoring spread and gauging the impact of interventions. Measuring diversity is also highly dependent on sampling, and in areas with high levels of genomic sequencing groups have used measurements such as the Shannon Index [113]. Higher-order measurements that weigh the major circulating lineages (eg, Simpson Index) may be less prone to bias from under-sampling. We have recently argued that measurements of effective circulating diversity (Hill numbers [115–117]) may be more reliable, and less biased, especially in under-sampled locations [75].

One other way that whole-genome sequencing has been helpful at the very local level is in contact tracing and confirming outbreaks. In this regard, whole-genome sequences can either rule in or rule out transmission events. In some reported instances, transmitted viruses are either completely identical at the nucleotide level or 1-3 SNP differences apart [118, 119]. To conclusively demonstrate transmission, it is necessary to compare the genomes in a putative cluster to other circulating genomes in the community and also establish epidemiological links. Studies that have looked at clusters in this way have generally been able to separate true transmission events from instances that may have an epidemiological link but different genomes.

BIOLOGICAL DIFFERENCES BETWEEN GENOMIC VARIANTS

Differences in Transmission

While the most clear biological difference between genomic variants is in transmission, even this property is unclear and riddled with problems of bias. Of the variants that have been proposed to have an increased rate of transmission, the D614G mutation in the spike protein has the most evidence behind it from epidemiological studies, phylodynamic modeling, structural analyses, and in vivo and in vitro comparative analyses [27, 120–122]. Other variants have been less well studied, with the N501Y mutation garnering, perhaps, the second-best support [79, 96, 123–125]. Both the B.1.1.7 and the B.1.351 variants, which have the N501Y mutation in their RBDs, have recently been shown to have enhanced affinity for the ACE2 protein [98]. It is also important to distinguish here between a holistic view of the viral lineage (genomic background) in general and the specific mutational variants at specific positions. For instance, it could be that N501Y in one genomic context significantly increases transmission whereas in another context it has little impact.

In addition, an important quality of SARS-CoV-2 transmission appears to be its variable (over-dispersed) infectivity rate and its tendency to occur in super spreader events [126]. It is unclear whether the propensity to sometimes have a much higher R0 is primarily based on the virus, the host, behavior, or environment; it is likely a complex combination of these factors.

Disease Severity

Hypotheses that some variants cause worse disease have been proposed at various times during the pandemic [127], but attempts to understand outcomes have been limited by available sequencing data, biased sampling, and changing epidemiology. Most notably, after initial reports suggested that B.1.1.7 lineage was no more virulent than other strains [100, 128–130], subsequent reports suggested that it may, indeed, be associated with worse outcomes [131–135] although increased severity has not been shown in some recent studies [79, 125]. One study linked higher viral load to increased morbidity [136], and B.1.1.7 has been consistently noted to have higher titers in clinical samples [125, 137]. Recently, there have been recent reports that B.1.351 and P.2 may also be associated with poorer outcomes [138, 139]. All of these studies suffer from the enormous problems of biased sampling and will need well-designed and controlled clinical studies combined with whole-genome sequencing to make more definitive statements.

In general, pediatric cases of COVID-19 are less severe [140, 141], and as such there have been only a few reported genome studies of sequences from patients under 21 [142–145]. One study found an association between genotype and severity [142].

Another presentation of SARS-CoV-2–related disease in the pediatric population has been the multisystem inflammatory syndrome in children (MIS-C), which is sometimes referred to as PIMS (pediatric inflammatory multisystem syndrome) [146–148]. This syndrome usually presents 4-6 weeks after infection with body-wide inflammatory signs and symptoms and, often, a cardiovascular component requiring hemodynamic support and treatment with steroids and intravenous immunoglobulin. Because MIS-C is a late presentation, the virus is often not present or found at very low levels in respiratory secretions. Therefore, whole-genome analysis has been difficult in this condition, and only a few reported sequences are available [149, 150]. Although the pathogenesis of MIS-C is not well understood, it has been suggested that components of the virus may act as superantigens to provoke a polyclonal T-cell expansion [151, 152] making it conceivable that viral variants may be more

or less likely to cause disease. The available sequences suggest that MIS-C associated viruses might be drawn from several distinct lineages that are representative of circulating virus without any genetic similarities that tie them together [150]. However, it is notable that the original experience with the virus in China did not detect any MIS-C, raising the possibility that this is presentation is linked to later forms of the virus. Much more viral sequencing will be required to test the hypothesis that some viral variants may be more likely to cause disease.

Vaccine Escape

In vitro studies of neutralizing antibodies from patients naturally infected with SARS-CoV-2 or immunized with different vaccines have clearly shown reductions in neutralizing antibody efficacy associated with specific variants. Most notable of these are mutations at the E484 position [130, 153, 154]. Mutations in naturally occurring variants at this position (E484K, E484Q) have been found in the B.1.351 (South African), P.1 and P.2 (Brazilian), and B.1.617 (Indian) lineages, with a small number of reported B.1.1.7 isolates bearing a mutation at this site [108]. Several studies have looked at neutralizing antibodies in each of these variants, and several have shown decreased neutralizing response compared to wild type [106, 155–157], however, the B.1.351 appears to have the largest decrease [102, 103]. Data like this suggest that specific variant positions may have distinct impacts in distinct lineages.

Despite significant reductions in neutralizing antibodies, many of the vaccines have been found to be effective, most encouragingly for severe disease, in places dominated by worrisome "vaccine escape" mutants [158–169]. The biggest drops in efficacy have been associated with the B.1.351 lineage but only in mild to moderate disease [167]. This protective effect is likely because of very good initial vaccine efficacy as well as retention of multiple other epitope targets, and potentially other immunological factors such as the T-cell response. Going forward we will need to develop quick assays for assessing correlates of protection that can be applied to emerging variants [170].

WHAT DO WE NEED TO DO GOING FORWARD?

Most evolutionary biology to date has been done in retrospect, and therefore the techniques are focused on deriving the maximum amount of data from things that have already happened. Likewise, epidemiological techniques are often centered around known diversity, and a pathogen that may not be changing its ability to spread or cause disease. In this instance, we need to reboot our toolbox and scale up our understanding of pathogen diversity. It will be absolutely critical moving forward to have, in place, systematic strategies for sampling, quickly analyzing, new data. It is instructive that patterns can change very rapidly. For instance, in Philadelphia between February and March our surveillance data saw a more than 400% increase in B.1.1.7. Our sequencing and analytical strategies need to be able to react in real time before the speeding train passes us by.

Notes

Authors contributions. Both authors contributed equally to writing, literature search, reviewing, and editing.

Financial support. P.J.P. and A.M.M. are supported by CDC200-2021-11102 from the Centers for Disease Control and Prevention (CDC) and funding from the Women's Committee of the Children's Hospital of Philadelphia (CHOP).

Potential conflicts of interest. Both authors: No potential conflicts of interest. Both authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet **2020**; 395(10223):497–506.
2. Allam Z. Chapter 1 - the first 50 days of COVID-19: a detailed chronological timeline and extensive review of literature documenting the pandemic. In: Allam Z, ed. Surveying the Covid-19 Pandemic and Its Implications. Amsterdam, Netherlands: Elsevier; **2020**:1–7.
3. WHO. Novel Coronavirus – Thailand (ex-China). Accessed May 27, 2021. https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON234.
4. WHO. Novel Coronavirus (2019-nCoV) Situation Report - 11. Accessed May 27, 2021. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4
5. Worobey M, Pekar J, Larsen BB, et al. The emergence of SARS-CoV-2 in Europe and North America. Science **2020**; 370(6516):564–70.
6. Nadeau SA, Vaughan TG, Scire J, Huisman JS, Stadler T. The origin and early spread of SARS-CoV-2 in Europe. Proc Natl Acad Sci U S A **2021**; 118(9):e2012008118.
7. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature **2020**; 579(7798):265–9.
8. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature **2020**; 579(7798):270–3.
9. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill **2020**; 25(3). pii=2000045. doi:10.2807/1560-7917.ES.2020.25.3.2000045
10. Thi Nhu Thao T, Labroussaa F, Ebert N, et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. Nature **2020**; 582:561–5.
11. Andersen KG, Rambaut A, Lipkin WI, et al. The proximal origin of SARS-CoV-2. Nat Med **2020**; 26:450–2.
12. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell **2020**; 181(2):271–80.e8.
13. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet **2020**; 395(10224):565–74.
14. Zhang T, Wu Q, Zhang Z. Pangolin homology associated with 2019-nCoV. bioRxiv **2020**:2020.02.19.950253.
15. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. Curr Biol **2020**; 30(7):1346–51.e2.
16. Lam TT, Jia N, Zhang YW, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature **2020**; 583:282–5.
17. Liu P, Jiang JZ, Wan XF, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog **2020**; 16(5):e1008421.
18. Xiao K, Zhai J, Feng Y, et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature **2020**; 583:286–9.
19. Wrobel AG, Benton DJ, Xu P, et al. Structure and binding properties of Pangolin-CoV spike glycoprotein inform the evolution of SARS-CoV-2. Nat Commun **2021**; 12:837.
20. Wrobel AG, Benton DJ, Xu P, et al. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. Nat Struct Mol Biol **2020**; 27:763–7.
21. Zhang S, Qiao S, Yu J, et al. Bat and pangolin coronavirus spike glycoprotein structures provide insights into SARS-CoV-2 evolution. Nat Commun **2021**; 12:1607.
22. Boni MF, Lemey P, Jiang X, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol **2020**; 5:1408–17.

23. Coutard B, Valle C, de Lamballerie X, et al. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res **2020**; 176:104742.
24. Johnson BA, Xie X, Bailey AL, et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. Nature **2021**; 591:293–9.
25. Peacock TP, Goldhill DH, Zhou J, et al. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. Nat Microbiol **2021**. doi:10.1038/s41564-021-00908-w
26. Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol Cell **2020**; 78(4):779–84.e5.
27. Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell **2020**; 182(4):812–27.e19.
28. Yang Q, Hughes TA, Kelkar A, et al. Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. eLife **2020**; 9:e61552.
29. Nao N, Yamagishi J, Miyamoto H, et al. Genetic predisposition to acquire a polybasic cleavage site for highly pathogenic avian influenza virus hemagglutinin. mBio **2017**; 8(1):e02298-16.
30. Menachery VD, Dinnon KH, Yount BL, et al. Trypsin treatment unlocks barrier for zoonotic bat coronavirus infection. J Virol **2020**; 94(5):e01774-19.
31. Alexander DJ, Brown IH. History of highly pathogenic avian influenza. Rev Sci Tech **2009**; 28:19–38.
32. Ito T, Goto H, Yamamoto E, et al. Generation of a highly pathogenic avian influenza A virus from an avirulent field isolate by passaging in chickens. J Virol **2001**; 75(9):4439–43.
33. Follis KE, York J, Nunberg JH. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. Virology **2006**; 350:358–69.
34. Watanabe Y, Allen JD, Wrapp D, et al. Site-specific glycan analysis of the SARS-CoV-2 spike. Science **2020**; 369:330–3.
35. Rasmussen AL. On the origins of SARS-CoV-2. Nat Med **2021**; 27:9.
36. Li X, Giorgi EE, Marichannegowda MH, et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci Adv **2020**; 6(27):eabb9153.
37. Liu SL, Saif LJ, Weiss SR, Su L. No credible evidence supporting claims of the laboratory engineering of SARS-CoV-2. Emerg Microbes Infect **2020**; 9:505–7.
38. Hakim MS. SARS-CoV-2, Covid-19, and the debunking of conspiracy theories [published online ahead of print February 14, **2021**]. Rev Med Virol doi:10.1002/rmv.2222.
39. Ghafari M, du Plessis L, Pybus O, Katzourakis A. Time dependence of SARS-CoV-2 substitution rates. Accessed May 27, 2021. https://virological.org/t/time-dependence-of-sars-cov-2-substitution-rates/542
40. Su YCF, Anderson DE, Young BE, et al. Discovery of a 382-nt deletion during the early evolution of SARS-CoV-2. bioRxiv **2020**:2020.03.11.987222.
41. MacLean OA, Orton RJ, Singer JB, Robertson DL. No evidence for distinct types in the evolution of SARS-CoV-2. Virus Evol **2020**; 6:veaa034.
42. Rambaut A. Phylogenetic analysis of nCoV-2019 genomes. Accessed May 27, 2021. https://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356
43. Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. Front Microbiol **2020**; 11:1800.
44. Domingo E, Perales C. Viral quasispecies. PLoS Genet **2019**; 15:e1008271.
45. Xu D, Zhang Z, Wang FS. SARS-associated coronavirus quasispecies in individual patients. N Engl J Med **2004**; 350:1366–7.
46. Park D, Huh HJ, Kim YJ, et al. Analysis of intrapatient heterogeneity uncovers the microevolution of Middle East respiratory syndrome coronavirus. Cold Spring Harb Mol Case Stud **2016**; 2:a001214.
47. Lau BT, Pavlichin D, Hooker AC, et al. Profiling SARS-CoV-2 mutation fingerprints that range from the viral pangenome to individual infection quasispecies. Genome Med **2021**; 13:62.
48. Jary A, Leducq V, Malet I, et al. Evolution of viral quasispecies during SARS-CoV-2 infection. Clin Microbiol Infect **2020**; 26(11):1560.e1–e4.
49. Capobianchi MR, Rueca M, Messina F, et al. Molecular characterization of SARS-CoV-2 from the first case of COVID-19 in Italy. Clin Microbiol Infect **2020**; 26:954–6.
50. Shen Z, Xiao Y, Kang L, et al. Genomic diversity of severe acute respiratory syndrome–coronavirus 2 in patients with coronavirus disease 2019. Clin Infect Dis **2020**; 71(15):713–20.
51. Karamitros T, Papadopoulou G, Bousali M, et al. SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies. J Clin Virol **2020**; 131:104585.
52. Andrés C, Garcia-Cehic D, Gregori J, et al. Naturally occurring SARS-CoV-2 gene deletions close to the spike S1/S2 cleavage site in the viral quasispecies of COVID19 patients. Emerg Microbes Infect **2020**; 9:1900–11.
53. Sun F, Wang X, Tan S, et al. SARS-CoV-2 Quasispecies provides insight into its genetic dynamics during infection. bioRxiv **2020**:2020.08.20.258376.
54. Siqueira JD, Goes LR, Alves BM, et al. SARS-CoV-2 genomic and quasispecies analyses in cancer patients reveal relaxed intrahost virus evolution. bioRxiv **2020**:2020.08.26.267831.
55. Wölfel R, Corman VM, Guggemos W, et al. Virological assessment of hospitalized patients with COVID-2019. Nature **2020**; 581(7809):465–9.
56. Popa A, Genger J-W, Nicholson MD, et al. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. Sci Transl Med **2020**; 12(573):eabe2555.
57. Xiao M, Liu X, Ji J, et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. Genome Med **2020**; 12:57.
58. Ko SH, Bayat Mokhtari E, Mudvari P, et al. High-throughput, single-copy sequencing reveals SARS-CoV-2 spike variants coincident with mounting humoral immunity during acute COVID-19. PLoS Pathog **2021**; 17:e1009431.
59. Müller NF, Kistler KE, Bedford T. Recombination patterns in coronaviruses. bioRxiv **2021**:2021.04.28.441806.
60. Han GZ. Pangolins harbor SARS-CoV-2-related coronaviruses. Trends Microbiol **2020**; 28:515–7.
61. Patiño-Galindo JÁ, Filip I, Chowdhury R, et al. Recombination and lineage-specific mutations linked to the emergence of SARS-CoV-2. bioRxiv **2021**:2020.02.10.942748.
62. Latinne A, Hu B, Olival KJ, et al. Origin and cross-species transmission of bat coronaviruses in China. Nat Commun **2020**; 11:4235.
63. Jackson B, Rambaut A, Pybus OG, et al. Recombinant SARS-CoV-2 genomes involving lineage B.1.1.7 in the UK. Accessed May 27, 2021. https://virological.org/t/recombinant-sars-cov-2-genomes-involving-lineage-b-1-1-7-in-the-uk/658
64. VanInsberghe D, Neish AS, Lowen AC, Koelle K. Recombinant SARS-CoV-2 genomes are currently circulating at low levels. bioRxiv **2021**:2020.08.05.238386.
65. Pybus OG. Pango Lineage Nomenclature: provisional rules for naming recombinant lineages. Accessed May 27, 2021. https://virological.org/t/pango-lineage-nomenclature-provisional-rules-for-naming-recombinant-lineages/657
66. Pedro N, Silva CN, Magalhães AC, et al. Dynamics of a dual SARS-CoV-2 lineage co-infection on a prolonged viral shedding COVID-19 case: insights into clinical severity and disease duration. Microorganisms **2021**; 9(2):300.
67. Francisco RDS Jr, Benites LF, Lamarca AP, et al. Pervasive transmission of E484K and emergence of VUI-NP13L with evidence of SARS-CoV-2 co-infection events by two different lineages in Rio Grande do Sul, Brazil. Virus Res **2021**; 296:198345.
68. Cherian S, Potdar V, Jadhav S, et al. Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. bioRxiv **2021**:2021.04.22.440932.
69. Martin DP, Weaver S, Tegally H, 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**:2021.02.23.21252268.
70. Gu H, Chen Q, Yang G, et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. Science **2020**; 369:1603–7.
71. Zhou H-Y, Ji C-Y, Fan H, et al. Convergent evolution of SARS-CoV-2 in human and animals. Protein Cell **2021**. doi:10.1007/s13238-021-00847-6
72. Rambaut A, Holmes EC, O'Toole A, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol **2020**; 5:1403–7.
73. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics **2018**; 34(23):4121–3.
74. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill **2017**; 22(13). pii=30494.
75. Moustafa AM, Planet PJ. Emerging SARS-CoV-2 diversity revealed by rapid whole genome sequence typing. bioRxiv **2020**. doi:10.1101/2020.12.28.424582
76. Cleemput S, Dumon W, Fonseca V, et al. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. Bioinformatics **2020**; 36(11):3552–5.
77. Maan H, Mbareche H, Raphenya AR, et al. Genotyping SARS-CoV-2 through an interactive web application. Lancet Digit Health **2020**; 2:e340–1.
78. Volz EM, Kosakovsky Pond SL, Ward MJ, et al. Phylodynamics of infectious disease epidemics. Genetics **2009**; 183:1421–30.
79. Davies NG, Abbott S, Barnard RC, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science **2021**; 372(6538):eabg3055.
80. Hu J, He C-L, Gao Q-Z, et al. D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity. bioRxiv **2020**:2020.06.20.161323.
81. Ozono S, Zhang Y, Ode H, et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. Nat Commun **2021**; 12:848.
82. Daniloski Z, Jordan TX, Ilmain JK, et al. The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. eLife **2021**; 10:e65365.

83. Yurkovetskiy L, Wang X, Pascal KE, et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. Cell **2020**; 183(3):739–51.e8.
84. Zhang L, Jackson CB, Mou H, et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nat Commun **2020**; 11:6013.
85. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. Science **2020**; 369(6501):297–301.
86. Volz E, Mishra S, Chand M, et al.; COVID-19 Genomics UK (COG-UK) Consortium. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature **2021**; 593:266–9.
87. Public Health England. Investigation of novel SARS-CoV-2 variant: variant of Concern 202012/01 (Technical briefing 5). **2021**. Accessed May 27, 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959426/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5.pdf
88. CDC. SARS-CoV-2 Variant Classifications and Definitions. Accessed June 17, 2021. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fcases-updates%2Fvariant-surveillance%2Fvariant-info.html
89. Rambaut A, Loman NJ, Pybus OG, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Accessed May 27, 2021. https://virological.org/t/563
90. Avanzato VA, Matson MJ, Seifert SN, et al. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with Cancer. Cell **2020**; 183(7):1901–12.e9.
91. Abbasi J. Researchers tie severe immunosuppression to chronic COVID-19 and virus variants. JAMA **2021**; 325:2033–5.
92. Choi B, Choudhary MC, Regan J, et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. N Engl J Med **2020**; 383:2291–3.
93. Kemp SA, Collier DA, Datir RP, et al.; CITIID-NIHR BioResource COVID-19 Collaboration; COVID-19 Genomics UK (COG-UK) Consortium. SARS-CoV-2 evolution during treatment of chronic infection. Nature **2021**; 592:277–82.
94. McCarthy KR, Rennick LJ, Nambulli S, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science **2021**; 371:1139–42.
95. Kemp SA, Harvey WT, Lytras S, Carabelli AM, Robertson DL, Gupta RK. Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion H69/V70. bioRxiv **2021**:2020.12.14.422555.
96. Ali F, Kasry A, Amin M. The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant. Med Drug Discov **2021**; 10:100086.
97. Luan B, Wang H, Huynh T. Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. FEBS Lett **2021**; 595:1454–61.
98. Ramanathan M, Ferguson ID, Miao W, Khavari PA. SARS-CoV-2 B.1.1.7 and B.1.351 spike variants bind human ACE2 with increased affinity. Lancet Infect Dis **2021**. doi:10.1016/S1473-3099(21)00262-0
99. O'Toole Á, Hill V, Pybus OG, et al.; COVID-19 Genomics UK (COG-UK) Consortium; Network for Genomic Surveillance in South Africa (NGS-SA); Brazil-UK CADDE Genomic Network; Swiss Viollier Sequencing Consortium; SEARCH Alliance San Diego; National Virus Reference Laboratory; SeqCOVID-Spain; Danish Covid-19 Genome Consortium (DCGC); Communicable Diseases Genomic Network (CDGN); Dutch National SARS-CoV-2 surveillance program; Division of Emerging Infectious Diseases (KDCA). Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2. Wellcome Open Res **2021**; 6:121.
100. Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. eLife **2020**; 9:e61312.
101. Wang P, Casner RG, Nair MS, et al. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. Cell Host Microbe **2021**; 29(5):747–51.e4.
102. Wang P, Nair MS, Liu L, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature **2021**; 593:130–5.
103. Zhou D, Dejnirattisai W, Supasa P, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell **2021**; 184:2348–2361.e6.
104. Garcia-Beltran WF, Lam EC, St Denis K, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell **2021**; 184(9):2372–2383.e9.
105. Wu K, Werner AP, Koch M, et al. Serum neutralizing activity elicited by mRNA-1273 vaccine. N Engl J Med **2021**; 384(15):1468–70.
106. Collier DA, De Marco A, Ferreira IATM, et al.; CITIID-NIHR BioResource COVID-19 Collaboration; COVID-19 Genomics UK (COG-UK) Consortium. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. Nature **2021**; 593:136–41.
107. Liu Y, Liu J, Xia H, et al. Neutralizing activity of BNT162b2-elicited serum. N Engl J Med **2021**; 384:1466–8.
108. Moustafa AM, Bianco C, Denu L, et al. Comparative analysis of emerging B.1.1.7+E484K SARS-CoV-2 isolates from Pennsylvania. bioRxiv **2021**:2021.04.21.440801. doi:10.1093/ofid/ofab300
109. Deng X, Garcia-Knight MA, Khalid MM, et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. Cell **2021**. doi:10.1016/j.cell.2021.04.025
110. Di Giacomo S, Mercatelli D, Rakhimov A, Giorgi FM. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K [published online ahead of print May 5, **2021**]. J Med Virol doi:10.1002/jmv.27062.
111. Public Health England. Risk assessment for SARS-CoV-2 variant: VOC-21APR-02 (B.1.617.2). **2021**. Accessed May 27, 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/990101/27_May_2021_Risk_assessment_for_SARS-CoV-2_variant_VOC-21APR-02__B.1.617.2_.pdf
112. Seemann T, Lane CR, Sherry NL, et al. Tracking the COVID-19 pandemic in Australia using genomics. Nat Commun **2020**; 11:4376.
113. du Plessis L, McCrone JT, Zarebski AE, et al.; COVID-19 Genomics UK (COG-UK) Consortium. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science **2021**; 371:708–12.
114. Zuckerman NS, Bucris E, Drori Y, et al. Genomic variation and epidemiology of SARS-CoV-2 importation and early circulation in Israel. PLoS One **2021**; 16:e0243265.
115. Hill MO. Diversity and evenness: a unifying notation and its consequences. Ecology **1973**; 54(2):427–32.
116. Jost L. Entropy and diversity. Oikos **2006**; 113(2):363–75.
117. Alberdi A, Gilbert MTP. A guide to the application of Hill numbers to DNA-based diversity analyses. Mol Ecol Resour **2019**; 19(4):804–17.
118. Böhmer MM, Buchholz U, Corman VM, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. Lancet Infect Dis **2020**; 20:920–8.
119. Günther T, Czech-Sioli M, Indenbirken D, et al. SARS-CoV-2 outbreak investigation in a German meat processing plant. EMBO Mol Med **2020**; 12:e13296.
120. Plante JA, Liu Y, Liu J, et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature **2021**; 592:116–21.
121. Hou YJ, Chiba S, Halfmann P, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science **2020**; 370:1464–8.
122. Zhang J, Cai Y, Xiao T, et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. Science **2021**; 372:525–30.
123. Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill **2021**; 26(1):2002106.
124. Liu Y, Liu J, Plante KS, et al. The N501Y spike substitution enhances SARS-CoV-2 transmission. bioRxiv **2021**:2021.03.08.434499.
125. Frampton D, Rampling T, Cross A, et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. Lancet Infect Dis **2021**. doi:10.1016/S1473-3099(21)00262-0
126. Lemieux JE, Siddle KJ, Shaw BM, et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science **2021**; 371(6529):eabe3261.
127. Voss JD, Skarzynski M, McAuley EM, et al. Variants in SARS-CoV-2 associated with mild or severe outcome. medRxiv **2020**:2020.12.01.20242149.
128. Wu K, Werner AP, Moliva JI, et al. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. bioRxiv **2021**:2021.01.25.427948.
129. Xie X, Zou J, Fontes-Garfias CR, et al. Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. bioRxiv **2021**:2021.01.07.425740.
130. Greaney AJ, Loes AN, Crawford KHD, 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 **2021**; 29(3):463–76.e6.
131. Horby P, Bell I, Breuer J, et al. NERVTAG: update note on B.1.1.7 severity. Paper prepared by the New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) on SARS-CoV-2 variant B.1.1.7, **2021**.
132. Davies NG, Jarvis CI, Edmunds WJ, et al.; CMMID COVID-19 Working Group. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature **2021**; 593:270–4.
133. Grint DJ, Wing K, Williamson E, et al. Case fatality risk of the SARS-CoV-2 variant of concern B.1.1.7 in England, 16 November to 5 February. Euro Surveill **2021**; 26(11):2100256.
134. Patone M, Thomas K, Hatch R, et al. Analysis of severe outcomes associated with the SARS-CoV-2 Variant of Concern 202012/01 in England

using ICNARC Case Mix Programme and QResearch databases. medRxiv **2021**:2021.03.11.21253364.
135. Bager P, Wohlfahrt J, Fonager J, et al. Increased Risk of Hospitalisation Associated with Infection with SARS-CoV-2 Lineage B.1.1.7 in Denmark. **2021**. Accessed May 27, 2021. https://ssrn.com/abstract=3792894
136. Fajnzylber J, Regan J, Coxen K, et al.; Massachusetts Consortium for Pathogen Readiness. SARS-CoV-2 viral load is associated with increased disease severity and mortality. Nat Commun **2020**; 11:5493.
137. Golubchik T, Lythgoe KA, Hall M, et al. Early analysis of a potential link between viral load and the N501Y mutation in the SARS-COV-2 spike protein. medRxiv **2021**:2021.01.12.20249080.
138. Funk T, Pharris A, Spiteri G, et al. Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. Euro Surveill **2021**; 26(16):2100348.
139. Pearson CAB, Russell TW, Davies NG, et al. Estimates of severity and transmissibility of novel South Africa SARS-CoV-2 variant 501Y.V2. **2021**. Accessed May 27, 2021. https://cmmid.github.io/topics/covid19/reports/sa-novel-variant/2021_01_11_Transmissibility_and_severity_of_501Y_V2_in_SA.pdf
140. Zhang C, Gu J, Chen Q, et al. Clinical and epidemiological characteristics of pediatric SARS-CoV-2 infections in China: a multicenter case series. PLoS Med **2020**; 17:e1003130.
141. Götzinger F, Santiago-García B, Noguera-Julián A, et al. COVID-19 in children and adolescents in Europe: a multinational, multicentre cohort study. Lancet Child Adolesc Health **2020**; 4(9):653–61.
142. Pandey U, Yee R, Shen L, et al. High prevalence of SARS-CoV-2 genetic variation and D614G mutation in pediatric patients with COVID-19. Open Forum Infect Dis **2021**; 8:ofaa551.
143. LoTempio J, Billings E, Draper K, et al. Novel SARS-CoV-2 spike variant identified through viral genome sequencing of the pediatric Washington D.C. COVID-19 outbreak. medRxiv **2021**:2021.02.08.21251344.
144. Moustafa AM, Otto W, Gai X, et al. Early pandemic molecular diversity of SARS-CoV-2 in children. medRxiv **2021**:2021.02.17.21251960.
145. Bard JD, Bootwalla M, Leber A, et al. Emergence of SARS-CoV-2 variants of concern in the pediatric population of the United States. medRxiv **2021**:2021.05.22.21257660.
146. Feldstein LR, Rose EB, Horwitz SM, et al.; Overcoming COVID-19 Investigators; CDC COVID-19 Response Team. Multisystem inflammatory syndrome in U.S. children and adolescents. N Engl J Med **2020**; 383:334–46.
147. Davies P, Evans C, Kanthimathinathan HK, et al. Intensive care admissions of children with paediatric inflammatory multisystem syndrome temporally associated with SARS-CoV-2 (PIMS-TS) in the UK: a multicentre observational study. Lancet Child Adolesc Health **2020**; 4:669–77.
148. Sancho-Shimizu V, Brodin P, Cobat A, et al. SARS-CoV-2-related MIS-C: a key to the viral and genetic causes of Kawasaki disease? J Exp Med **2021**; 218(6): e20210446.
149. Parsons E, Timlin M, Starr C, et al. Multisystem inflammatory syndrome in children in February 2020 and implications of genomic sequencing for SARS-CoV-2. J Pediatric Infect Dis Soc **2021**; 10(5):695–7.
150. Pang J, Boshier FAT, Alders N, Dixon G, Breuer J. SARS-CoV-2 polymorphisms and multisystem inflammatory syndrome in children. Pediatrics **2020**; 146(6):e2020019844.
151. Cheng MH, Zhang S, Porritt RA, et al. Superantigenic character of an insert unique to SARS-CoV-2 spike supported by skewed TCR repertoire in patients with hyperinflammation. Proc Natl Acad Sci U S A **2020**; 117:25254–62.
152. Moreews M, Le Gouge K, Khaldi-Plassart S, et al. Polyclonal expansion of TCR Vbeta 21.3^{+} CD4^{+} and CD8^{+} T cells is a hallmark of multisystem inflammatory syndrome in children. Sci Immunol **2021**; 6(59):eabh1516.
153. Jangra S, Ye C, Rathnasinghe R, et al. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. Lancet Microbe **2021**. doi:10.1016/S2666-5247(21)00068-9
154. Liu Z, VanBlargan LA, Bloyet L-M, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe **2021**; 29(3):477–88.e4.
155. Abdool Karim SS, de Oliveira T. New SARS-CoV-2 variants—clinical, public health, and vaccine implications. N Engl J Med **2021**; 384(19):1866–8.
156. Cele S, Gazy I, Jackson L, et al.; Network for Genomic Surveillance in South Africa; COMMIT-KZN Team. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. Nature **2021**; 593:142–6.
157. Dejnirattisai W, Zhou D, Supasa P, et al. Antibody evasion by the P.1 strain of SARS-CoV-2. Cell **2021**; 184(11):2939–54.e9.
158. Hall VJ, Foulkes S, Saei A, et al.; SIREN Study Group. COVID-19 vaccine coverage in health-care workers in England and effectiveness of BNT162b2 mRNA vaccine against infection (SIREN): a prospective, multicentre, cohort study. Lancet **2021**; 397:1725–35.
159. Dagan N, Barda N, Kepten E, et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. N Engl J Med **2021**; 384:1412–23.
160. Goldberg Y, Mandel M, Woodbridge Y, et al. Protection of previous SARS-CoV-2 infection is similar to that of BNT162b2 vaccine protection: a three-month nationwide experience from Israel. medRxiv **2021**:2021.04.20.21255670.
161. Haas EJ, Angulo FJ, McLaughlin JM, et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. Lancet **2021**; 397:1819–29.
162. Fabiani M, Ramigni M, Gobbetto V, Mateo-Urdiales A, Pezzotti P, Piovesan C. Effectiveness of the Comirnaty (BNT162b2, BioNTech/Pfizer) vaccine in preventing SARS-CoV-2 infection among healthcare workers, Treviso province, Veneto region, Italy, 27 December 2020 to 24 March 2021. Euro Surveill **2021**; 26(17):2100420.
163. Moustsen-Helms IR, Emborg H-D, Nielsen J, et al. Vaccine effectiveness after 1st and 2nd dose of the BNT162b2 mRNA Covid-19 Vaccine in long-term care facility residents and healthcare workers – a Danish cohort study. medRxiv **2021**:2021.03.08.21252200.
164. FDA. Vaccines and Related Biological Products Advisory Committee Meeting February 26, 2021. **2021**. Accessed May 27, 2021. https://www.fda.gov/media/146217/download
165. Novavax. Novavax COVID-19 Vaccine Demonstrates 89.3% Efficacy in UK Phase 3 Trial. **2021**. Accessed May 27, 2021. https://ir.novavax.com/news-releases/news-release-details/novavax-covid-19-vaccine-demonstrates-893-efficacy-uk-phase-3
166. Shinde V, Bhikha S, Hoosain Z, et al.; 2019nCoV-501 Study Group. Efficacy of NVX-CoV2373 Covid-19 vaccine against the B.1.351 Variant. N Engl J Med **2021**; 384:1899–909.
167. Madhi SA, Baillie V, Cutland CL, et al.; NGS-SA Group; Wits-VIDA COVID Group. Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B.1.351 variant. N Engl J Med **2021**; 384:1885–98.
168. Emary KRW, Golubchik T, Aley PK, et al.; COVID-19 Genomics UK Consortium; AMPHEUS Project; Oxford COVID-19 Vaccine Trial Group. Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory analysis of a randomised controlled trial. Lancet **2021**; 397:1351–62.
169. Bernal JL, Andrews N, Gower C, et al. Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. medRxiv **2021**:2021.05.22.21257658.
170. Earle KA, Ambrosino DM, Fiore-Gartland A, et al. Evidence for antibody as a protective correlate for COVID-19 vaccines. medRxiv **2021**:2021.03.17.20200246.