The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic

Sarah P. Otto*, Troy Daya, Julien Arinob, Caroline Colinab, Jonathan Dushoffd, Michael Lid, Samir Mechial, Gary Van Domselaara,e, Jianhong Wu,f, David J.D. Earnp, and Nicholas H. Ogdenq

1Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
2Department of Mathematics and Statistics, Department of Biology, Queen’s University, Kingston, ON K7L 3N6, Canada
3Department of Mathematics and Data Science Nexus, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
4Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
5Department of Biology and M.G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON L8S 4L8, Canada
6Public Health Risk Sciences Division, National Microbiology Laboratory, Public Health Agency of Canada, Guelph, ON N1G 3W4, Canada
7Public Health Risk Sciences Division, National Microbiology Laboratory, Public Health Agency of Canada, St. Hyacinthe, QC J2S 2M2, Canada
8National Microbiology Laboratory – Public Health Agency of Canada, Winnipeg, MB R3E 3R2, Canada
9Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, MB R3E 0J9, Canada
10Laboratory for Industrial and Applied Mathematics, York University, Toronto, ON M3J 1P3, Canada
11Department of Mathematics and Statistics and M.G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON L8S 4K1, Canada

*Correspondence: otto@zoology.ubc.ca
https://doi.org/10.1016/j.cub.2021.06.049

SUMMARY

One year into the global COVID-19 pandemic, the focus of attention has shifted to the emergence and spread of SARS-CoV-2 variants of concern (VOCs). After nearly a year of the pandemic with little evolutionary change affecting human health, several variants have now been shown to have substantial detrimental effects on transmission and severity of the virus. Public health officials, medical practitioners, scientists, and the broader community have since been scrambling to understand what these variants mean for diagnosis, treatment, and the control of the pandemic through nonpharmaceutical interventions and vaccines. Here we explore the evolutionary processes that are involved in the emergence of new variants, what we can expect in terms of the future emergence of VOCs, and what we can do to minimise their impact.

Introduction

The December 2020 announcement by Public Health England that the variant B.1.1.7 exhibited a large number of mutations and a significant increase in transmission (50–100%2) rattled the world and served as a wake-up call on the importance of VOCs. Until then, there was little evidence that any mutations to the RNA genome of SARS-CoV-2 substantially increased viral fitness3. Within weeks, additional VOCs were reported with similar characteristics: greater than expected numbers of mutations and signatures of enhanced transmission (for example, B.1.351 in South Africa4 and P.1 in Brazil5).

The genomes of all viruses accumulate mutations over time. However, the pace of mutation accumulation and the consequences for transmission and disease in the host population depend on a range of factors, including the mutation rate and the impacts of mutation on viral dynamics within and between individual hosts. Together these factors determine the emergence and spread of viral variants and the evolution of pandemics. The genomes of RNA viruses accumulate mutations over time. However, the pace of mutation accumulation and the consequences for transmission and disease in the host population depend on a range of factors, including the mutation rate and the impacts of mutation on viral dynamics within and between individual hosts. Together these factors determine the emergence and spread of viral variants and the evolution of pandemics. The genomes of RNA viruses are particularly prone to mutation6. But SARS-CoV-2, like related coronaviruses, encodes a proof-reading domain (ExoN) that reduces its mutation rate relative to RNA viruses that do not (such as influenza, HIV, and hepatitis C viruses). The pace of change of the SARS-CoV-2 genome has thus been estimated at 1.87 × 10⁻⁶ nucleotide substitutions per site per day, roughly 5-fold lower than influenza A/H3N2 (10.9 × 10⁻⁶ nucleotide substitutions per site per day; https://nextstrain.org; accessed 22 May 2021). Thus, across the ~30,000 base-pair genome of SARS-CoV-2, approximately 20 genetic changes occur per year within a lineage.

Not all mutations that arise will persist long enough to be included in this estimate of the molecular clock of SARS-CoV-2. Assuming that synonymous mutations are largely neutral while non-synonymous mutations persist with probability ω, Wang et al.9 used codon-based likelihood methods to estimate ω = 0.56 among human SARS-CoV-2. Thus, roughly half of the mutations that alter amino acids are quickly lost. Given the proportion of sites subject to synonymous (pS = 0.22) and non-synonymous (pN = 0.78) mutations, we can thus back-calculate the rate at which mutations must have occurred, before selection acted, to be 2.84 × 10⁻⁶ (that is, 1.87 × 10⁻⁶/(pS + ω pN)). This correction is slight because selection has had little time to eliminate anything but the most deleterious mutations accumulated over the pandemic’s short time frame. Over longer evolutionary time frames, such as that observed between SARS-CoV-2 and the most closely related bat sequences, Wang et al.9 find that non-synonymous mutations are 14 times less likely to persist (ω = 0.039) than what we see today among the virus in humans, as selection has slowly eliminated weakly deleterious mutations.
These rapid changes are faster than expected by drift in a large viral population within a patient, over the course of days. Studies have documented rapid changes to the composition of infections. By sequencing the virus at multiple time points, these processes have been best documented in immunocompromised individuals. For SARS-CoV-2, these within-individual evolutionary processes have been best documented in immunocompromised patients\(^{14}\). These patients maintain high viral loads over prolonged periods of time, allowing more opportunities for viral replication and selection and leading to elevated substitution rates (see elevated branch lengths in, for example, references\(^ {11,13}\), which reflect a larger than expected number of substitutions). By sequencing the virus at multiple time points, these studies have documented rapid changes to the composition of the viral population within a patient, over the course of days. These rapid changes are faster than expected by drift in a large population and point to natural selection for the virus to better replicate or to evade a weakened immune system, as well as the intensive antibody therapies received by these patients\(^ {15–13}\). A disproportionate fraction of mutations in many of these patients are clustered in the gene encoding Spike, the protein that juts out of the virus and binds to the ACE2 receptor that allows entry into host cells: for example, 57% of 15 changes in the final sample reported by Choi et al.\(^ {11}\) and 33% of 9 changes in the final nose and throat sample reported by Kemp et al.\(^ {12}\), compared to an expectation of only 13% based on the length of Spike. Many of these mutations were found in the receptor binding domain, which is essential for host entry. This concentration in Spike may reflect, in part, relaxed selective constraints within immunocompromised individuals, but many of these mutations are non-synonymous (for example, all eight mutations in Spike reported by Choi et al.\(^ {11}\)), consistent with selection favouring changes to the Spike protein. Furthermore, several mutations have arisen in parallel in different patients, suggesting selection for those changes\(^ {15}\). For example, McCarthy et al.\(^ {14}\) report four independent cases of patients bearing deletions in the same amino-terminal domain of Spike, which is important for host entry and antibody evasion. These non-random patterns of genomic changes suggest that selective pressures, alongside mutation, strongly shape viral evolution within immunocompromised individuals. Immunocompromised patients should be protected from COVID-19 by prioritizing their contacts for vaccination, and any patient with a prolonged infection should be treated with great care as a potential source of new variants.

The changes observed in the three main VOCs to date (B.1.1.7, P.1, B.1.351) echo the evolutionary changes observed in immunocompromised individuals, with a larger than expected number of mutations that are non-randomly distributed across the genome. This pattern led Public Health England\(^ {10}\) to suggest that B.1.1.7 might have originated in an immunocompromised individual. Many mutations in these VOCs are also concentrated in Spike.
Table 1. SARS-CoV-2 variants of concern and variants of interest.

<table>
<thead>
<tr>
<th>Pango lineage</th>
<th>Nextstrain clade</th>
<th>First detection location</th>
<th>VOI or VOC designation</th>
<th>WHO designation</th>
<th>Mutations of interest on S protein</th>
<th>Observed clinical effect</th>
<th>Transmissibility</th>
<th>Virulence</th>
<th>Antigenicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.1.1.7</td>
<td>20I/501Y/V1</td>
<td>United Kingdom</td>
<td>VOC Alpha</td>
<td>N501Y, E148K*, P681H, D614G</td>
<td></td>
<td>50–100% higher^2</td>
<td>39–72% more lethal^1</td>
<td>No impact on NBA</td>
<td>Minimal impact on NBS^a</td>
</tr>
<tr>
<td>B.1.351</td>
<td>20H/501Y/V2</td>
<td>South Africa</td>
<td>VOC Beta</td>
<td>N501Y, K417N, E484K, D614G</td>
<td></td>
<td>20–133% higher^2</td>
<td>–</td>
<td>Moderately reduced NBA^a</td>
<td>Reduced NBS^b</td>
</tr>
<tr>
<td>P.1</td>
<td>20J/501Y/V3</td>
<td>Brazil/Japan</td>
<td>VOC Gamma</td>
<td>N501Y, K417T, E484K, D614G</td>
<td></td>
<td>70–140% higher, evades immunity 21–46% more^3</td>
<td>20–90% more lethal^2</td>
<td>Moderately reduced NBA^a</td>
<td>Reduced NBS^b</td>
</tr>
<tr>
<td>B.1.427 and B.1.429</td>
<td>21C/S:452R</td>
<td>United States (California)</td>
<td>VOI Epsilon</td>
<td>L452R, D614G</td>
<td></td>
<td>18–22% higher^3</td>
<td>–</td>
<td>Moderately reduced NBA^a</td>
<td>Reduced NBS^b</td>
</tr>
<tr>
<td>B.1.526</td>
<td>21F</td>
<td>United States (New York)</td>
<td>VOI Iota</td>
<td>L5F*, T95I, D235G, S477N*, E484K, D614G, A701V*</td>
<td></td>
<td>–</td>
<td>–</td>
<td>Potentially reduced NBA^a</td>
<td>Reduced NBS^b</td>
</tr>
<tr>
<td>B.1.617.1</td>
<td>21B/S:154K</td>
<td>India</td>
<td>VOI Kappa</td>
<td>(T95I), G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H</td>
<td>Secondary attack rates similar to B.1.1.7</td>
<td>–</td>
<td>Potentially reduced NBA^a</td>
<td>Potentially reduced NBS^b</td>
<td></td>
</tr>
<tr>
<td>B.1.617.2</td>
<td>21A/S:478K</td>
<td>India</td>
<td>VOI Delta</td>
<td>T19R, G142D*, 156del, 157del, R158G, L452R, T478K, D614G, P681R, D950N</td>
<td>Secondary attack rates higher than B.1.1.7; household transmission 64% higher than B.1.1.7 (26–113% higher)^5</td>
<td>–</td>
<td>Potentially reduced NBA^a</td>
<td>Potentially reduced NBS^b</td>
<td></td>
</tr>
<tr>
<td>P.2</td>
<td>20J</td>
<td>Brazil</td>
<td>VOI Zeta</td>
<td>E484K, D614G, V1176F</td>
<td>–</td>
<td>–</td>
<td>Potentially reduced NBA^a</td>
<td>Reduced NBS^b</td>
<td></td>
</tr>
</tbody>
</table>

Here we have used the variant classification of the Centre for Disease Control of the United States and modified from the CDC data table, which lists additional mutations and information about clinical effects*. Nextstrain clade names start with the year of origin and distinguish lineages that reach a global frequency of 20%, so the name of a lineage can change if it increases in frequency. NBA: Neutralization by antibodies (monoclonal antibodies in therapeutic use); NBS: Neutralization by convalescent and/or post-vaccination sera (variant relative to non-variant). VOI and VOC designations vary over time (e.g. B.1.427 and B.1.429 were previously designated as VOC by the CDC^a, accessed 13 June 2021) and by country (e.g. Canada designates the entire B.1.617 clade as a VOC^b) because of different evaluations of the existing evidence.

*Mutation detected in some sequences within the lineage. Ranges give 95% confidence intervals or credible intervals, except for the transmission rate of B.1.1.7 (a consensus estimate across models) and for P.1 (50% Bayesian credible intervals).

^ahttps://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html; *moderate* includes modest decreases and/or cases where alternative antibody treatments remain available.

suggest that few viruses found subsequent infections; for example, only 1–8 virions. As a result, there is little genetic diversity, allowing little opportunity for evolutionary change within typical infections. Instead, drift is likely to dominate which viruses cause new infections, with one main exception: the elimination of strongly deleterious mutations. Indeed, Lythgoe et al. report a similar reduction in the number of nonsynonymous mutations observed within individuals as estimated from different individuals, suggesting that the elimination of strongly deleterious mutations primarily occurs within individual hosts.

Elevated mutation rates can also be caused by genetic changes in SARS-CoV-2 that increase the error rate during viral replication. Takada et al. showed that mutation P203L in NSP14 alters the ExoN proofreading domain and nearly doubles the mutation rate. Such “mutator” lineages could also generate variants with rare combinations of new mutations, but to date, these mutators have not risen in frequency, possibly because of the higher numbers of deleterious mutations that result.

Another potential source of novel variants within individual infections is recombination. There is debate about the extent of recombination in SARS-CoV-2. In part because detecting recombination is challenging when there is relatively little genomic variation. All circulating SARS-CoV-2 viruses are similar and closely related by descent from the virus that first infected humans in late 2019. Furthermore, artefacts generated during genomic sequencing and assembly can mimic recombination and may explain some of the apparent recombination events. That said, a recent analysis provides compelling evidence that sampled genomes from the UK arose by recombination during a period when B.1.1.7 and non-variant lineages were both prevalent. Recombination is a concern because it generates new combinations, potentially bringing together components of different lineages in a way that benefits the virus, although no fitness advantage has been detected for the recombinants observed to date.

Processes occurring among individuals

For those SARS-CoV-2 mutations that do make it out of the body and cause new infections, the probability that the variant becomes established depends on its transmissibility in the population in which it emerges, as well as the nature and extent of contacts among individuals and chance events. For many infectious diseases, a small fraction of individuals tends to be responsible for a large fraction of the transmission events (the so-called ‘80/20 rule’, in which 80% of new infections are thought to be caused by 20% of cases). This pattern has also been observed for SARS-COV-2 and is sometimes referred to as ‘overtaxis’, meaning that the variation in the number of new infections generated by different individuals is larger than expected (typically with reference to a Poisson distribution).

Overdispersion can arise through various processes, and the details have a strong effect on the probability that a new variant establishes within a population (by established, we mean that the variant is present in enough active cases that it is very unlikely to be lost by chance). If overdispersion arises because individuals vary in infectivity, with some cases transmitting much less often and others much more often than expected (for example, ‘superspreaders’), then this variability tends to decrease the probability of establishment of a variant (Figure 1, compare black and blue).
The key difference between the two cases is that when overdispersion occurs because some individuals have many fewer contacts or much lower activity levels and others many more, this tends to increase the probability of establishment (Figure 1, red curve compared to a Poisson distributed number of contacts in black).

The key difference between the two cases is that when overdispersion is due to chance events affecting the course of infection, then there is no relationship between the likelihood of an individual acquiring a variant infection and that individual’s propensity to transmit to others. As a result, this simply introduces additional noise into the system, decreasing the probability of establishment (blue curve). By contrast, with overdispersion due to contact or activity heterogeneity (red curve), variants are more likely to infect highly connected and active individuals who also have a higher propensity to transmit by virtue of their high connectivity. This process, whereby the most active individuals contribute disproportionately to transmission, also explains why targeting vaccines and nonpharmaceutical inter-views to essential workers is particularly effective at reducing the reproductive number (t). Even for those who do have symptoms, roughly half of the transmission events are estimated to occur before individuals exhibit symptoms (‘pre-symptomatic’ transmission), at least in areas where individuals are encouraged and able to self-isolate40. Most symptomatic individuals recover from COVID-19, but even so the infection fatality rate is high at about 0.68% (95% CI: 0.53–0.82%) globally41. In many infectious diseases, direct selection is expected to reduce virulence42. However, because severe symptoms and mortality from COVID-19 typically occur weeks after peak viral load, direct selection against virulence is expected to be extremely weak41–43. By contrast, selection strongly favours an increased transmission rate in the early stages of a pandemic, when the abundance of susceptible hosts is high44–48. These general predictions are also borne out in SARS-CoV-2-specific models, which additionally show that selection favours shorter latent periods prior to infectivity, higher infectiousness of asymptomatic individuals, and prolonged infectious periods42–44.

Although the death rate caused by SARS-CoV-2 may be under relatively weak direct selection, mutations that affect transmission rate or other disease attributes can have correlated effects on mortality. Variants with even a moderately enhanced transmission rate can readily spread during a pandemic, whether they increase or decrease death rates. Thus, the evolution of virulence depends strongly on the nature of the mutations that arise. For example, if increased transmission rate is due to a higher viral load, more severe illness may result. Alternatively, if mutations cause milder symptoms and so raise activity levels of infected individuals, they may increase transmissibility but reduce severity and death rates.

Unfortunately, of these possibilities, B.1.1.7 is both more transmissible and more virulent41,42. Based on structural modeling, the changes in the Spike protein in B.1.1.7 are predicted to cause conformational changes that allow the viruses to bind more easily to cells in the respiratory tract of an uninfected person43 (and so a lower infectious dose might be needed). Data also suggest that infected individuals have a slightly higher viral load and longer infectious period45. However, for B.1.1.7 at least, a shorter serial interval does not explain the observed spread in the UK46. The increased transmission has been accompanied by increased virulence, with an estimated 64% higher mortality rate (95% CI: 32–104%)42,44. Currently unknown is which of the mutations in B.1.1.7 are responsible for the observed changes in transmission and virulence, which are selectively favoured, and which have risen by hitchhiking despite being neutral or even deleterious.

While public health measures have reduced the reproductive number ($R_0$) of non-variant SARS-CoV-2 to below 1 in many countries, these measures were not initially strong enough to control the spread of B.1.1.7, allowing it to become the predominant variant in many countries within months. Where B.1.1.7 has established (say at 5–10% frequency), the subsequent rise in frequency has been fast. In the UK, the odds that a case was caused by B.1.1.7 (that is, the frequency of B.1.1.7 divided by the frequency of non-variants) doubled every 1–2 weeks depending on the region (Figure 2). The average reproductive number is expected to increase alongside the rise in variant frequency, requiring even stronger public health measures to control the disease. This recent history is now being
repeated with B.1.617.2, which is estimated to be 64% more transmissible within households than B.1.1.7 (Table 1).

**Processes occurring among species**

Predicting the course of SARS-CoV-2 evolution is further complicated by the broad potential host range and known transmissibility to animals. Were transmission in domesticated and/or wild animals to become an important reservoir for SARS-CoV-2, then selection for greater transmission in these animals could lead to different disease attributes. For example, evolution of transmissibility of West Nile virus in American robins (a key reservoir host) resulted in increased virulence to crows. For SARS-CoV-2, a host jump in Denmark from humans to farmed mink and back to humans led to a highly mutated variant (Cluster 5), which was less easily neutralized by antibodies from convalescent sera. Cluster 5 initially reached a high frequency (30%) in communities near affected farms, but through massive control efforts, including the culling of 3–4 million mink, appears (30%) in communities near affected farms, but through massive control efforts, including the culling of 3–4 million mink, appears to have been eradicated. As concluded by Larsen et al., keeping susceptible animals in such large and dense populations poses a threat to human health by presenting an opportunity for viral adaptation.

**Future evolution**

What types of variants are we likely to see over the next year? Although it is impossible to answer this question precisely, evolutionary models can help us explore the possibilities.

As long as circulating strains cannot infect recovered or immunized individuals, then models predict that, as the number of susceptible individuals falls, selection for increased transmissibility will become weaker, with selection increasingly favouring prolonged infectious periods.

However, as the proportion of the population that is immunologically naive shrinks through infection and immunization, selection will also increasingly favour variants that are partially or wholly able to overcome the immunity of previously protected individuals (Box 2). In some countries, particularly with limited medical and public-health resources, estimates suggest that most of the population may now have been infected, becoming naturally immune and increasing the relative fitness of such immune-escape variants.

The evolution of resistance to vaccines typically occurs more slowly and is observed less often than the evolution of resistance to therapeutics agents like antimicrobials. Both natural immunity and vaccine-induced immunity generally result in a broad range of antibodies and T cells that recognize distinct parts of the virus, particularly Spike and nucleoproteins. Even though current mRNA vaccines only encode Spike proteins, a range of antibody and T-cell responses are produced, recognizing different parts (‘epitopes’) of the protein, making it less likely that single mutational changes will substantially reduce protection. Even with the multiple mutations observed in the currently circulating VOCs, substantial natural and vaccine-induced antibody responses, and especially T-cell responses, are observed against the variants tested (B.1.1.7, B.1.351, and P.1). Importantly, a mutant virus that escapes the immune response of one individual may not escape the immune response of another.

Although full escape from vaccine-induced immunity in SARS-CoV-2 has not yet been documented, variants displaying partial escape are already here. P.1, B.1.617.2, and especially B.1.351 evade monoclonal antibody therapies, at least for some antibodies (Table 1). Many also exhibit weaker inhibition by neutralizing antibodies in serum samples from vaccinated individuals, requiring 6.4-fold higher titers of antibodies to neutralize B.1.351 and 3.5-fold to neutralize P.1 after two doses of Moderna (mRNA-1273). Real-world estimates also find slightly lower vaccine effectiveness against known infections by variants B.1.351 (Pfizer BNT162b2) and B.1.617.2 (Pfizer BNT162b2 and AstraZeneca ChAdOx1). Nevertheless, vaccine effectiveness increases with the second dose and/or over time, reaching high levels of protection against infection (>75%) and almost complete protection against severe disease.

Over the long term, the accumulation of multiple mutations in SARS-CoV-2 will eventually reduce the efficacy of immune responses in a large proportion of vaccinated and naturally immune people (Box 2). How long might we expect immunity to last? In another coronavirus (229E), antibodies within blood collected in the 1980s were shown to decline in neutralizing ability after a decade of evolutionary changes (‘antigenic drift’), suggesting years of immunity. But with so many active COVID-19 cases, further bursts of mutations, as seen in immunocompromised patients and via passage through animals, should be expected, accelerating evolutionary change. Altogether, it seems plausible that boosters may be needed within the next year or two, with longer gaps between boosters once case numbers fall globally. Fortunately, updating mRNA vaccines to target new variants is relatively straightforward, and clinical studies are already underway for variant-specific vaccines. For example, Moderna recently reported on phase 2 trials with mRNA-1273.351, a vaccine that matches changes in Spike observed in the B.1.351 variant. That said, the results showed very modest increases in antibodies targeting B.1.351 following two doses of the standard Moderna vaccine with mRNA-1273.351, suggesting that variant-specific vaccines may primarily boost previous immune responses rather than convey new immune responses to the variant epitopes. Targeting escape variants with vaccines that target other proteins, like the nucleocapsid, might be an effective alternative strategy.

Even without processes that make multiple changes more likely, selection can favour the accumulation of mutations one at a time that together reduce detection by the immune system and increase the risk of reinfection. The gradual build-up of escape mutants is particularly likely in regions where there are both high case numbers and high numbers of immunized individuals (Box 2).

In summary, the emergence of new variants should be expected, initially driven by selection for greater transmissibility and a longer duration of infectivity and then increasingly shaped by selection to evade immune responses, enabling transmission to vaccinated and naturally immunized people. Although there currently appears to be only weak selection on virulence, virulence may increase (or decrease) if coupled by mutation with these other features (as a pleiotropic side effect or through linkage). The further emergence of variants with multiple genetic changes, due to infections in immunocompromised individuals, mutators, and/or recombination, is expected and should be closely monitored.
Box 2. Escape mutations.

Each infection carries the risk of giving rise to a mutation that evades resistance in individuals who have specific immunity due to either prior infection or vaccination. To illustrate the main factors determining this risk, we simplify SARS-CoV-2 dynamics and consider a population consisting only of S susceptible individuals, I infected, and R resistant individuals, ignoring case heterogeneity (presymptomatic, asymptomatic, vaccination type and dose, etc.; see reference\(^67\) for a more detailed model focused on within-individual adaptation). Although these variables all change over time, we focus on a region with a small current fraction of active infections, so that S and R can be treated as constant over the short term, allowing us to focus on the dynamics of the infectious class I and escape variants I*.

Before escape variants are present, new cases arise at a rate \( \beta S I \) per day and are cleared at rate \( k \) (Box 2 figure, below). Each new infection has a chance \( m \) of giving rise to an escape variant by mutation. Because we are considering the first such mutation, the ability of the variant to infect resistant individuals is only partial, reducing the transmission rate to resistant individuals from \( \beta \) to \( p \beta \).

At this snapshot in the epidemic, the number of standard cases and the number of escape-variant cases are expected to change at rates:

\[
\frac{dI}{dt} = \left(1 - \mu\right) \frac{\beta S I}{C_0} - k I \tag{1}
\]

\[
\frac{dI^*}{dt} = \mu \frac{\beta S I}{C_0} + \frac{p \beta R I^*}{C_3} - k I^* \tag{2}
\]

The chance that an escape mutation appears on any particular day is thus proportional to \( (\mu \beta S I) \). While cases remain roughly constant in number, the time until the first escape mutation appears is approximately exponentially distributed with a mean of \( 1/(\mu \beta S I) \) days. The waiting time is thus shorter in locations with ineffective control measures (higher \( \beta \)), many circulating cases (higher \( I \)), and many remaining susceptibles (higher \( S \)).

Once an escape mutant has appeared, \textit{equation (2)} indicates that the mutant strain will increase in frequency at a faster rate than the original strain, raising the reproductive number for the new variant from \( R_t = (\beta S)/k \) to \( R_t = (\beta S + p \beta R)/k \). Preventing the selective accumulation of partially resistant mutations thus requires reducing contacts between cases and resistant individuals \( R I^* \), boosting resistance where possible by vaccinating naturally infected individuals and completing recommended vaccine doses (reducing \( p \)), and persisting with public health measures that reduce transmission in general (reducing \( \beta \)).

This basic model also clarifies the challenges inherent in deciding whether to spread available vaccines to more people in single doses or to complete two-dose regimes. Vaccinating more people with a single dose can decrease the number of infections circulating in a region (reducing \( I \) and \( I^* \)) and reduce the reproductive number of all variants (by reducing \( S \)), but it also has the potential to increase the probability of onward transmission to partially resistant individuals (increasing \( p R \) and the selective advantage of escape variants). The right policy thus depends on resistance after one versus two doses and the impact on the number of severe cases.

\[\text{A simplified model of the evolution of the first step escape mutant (}I^*\text{) using an SIR model framework.}\]

The transmission rate \( \beta \) to susceptible individuals and the clearance rate \( k \) are assumed to be equal for both variants. Only the escape mutant is partially able to infect resistant individuals (at a rate reduced by a factor \( p \) relative to the rate of infecting susceptible individuals). The model considers only the first step in the accumulation of mutations that increase the ability to reinfect otherwise resistant individuals but could be extended to model subsequent steps by mutation.
Effective ongoing surveillance also provides information about evolutionary change during the pandemic. Understanding the impact of these evolutionary changes can be important for guiding control measures and for solidifying public support for these measures. For example, as in many regions, B.1.1.7 recently increased in prevalence in Ontario Canada (Figure 3). During the initial spread (from January–February 2021), however, the total new daily case count steadily decreased, leading many people to believe that the pandemic was under control. But this steady decrease in total case count masked the fact that control measures at the time were insufficient to control B.1.1.7. As a result, once the prevalence of B.1.1.7 reached a high enough value, it began to dominate the total case count, causing the overall daily case count to start rising once again (Figure 3).

Although models predicted this evolutionary dynamic and the subsequent spike in cases based on the selective differences observed in the UK (Figure 4), accurate surveillance data for B.1.1.7 were not readily available to demonstrate unequivocally what was happening. Consequently, it was challenging to galvanize support for enhancing control measures proactively until the total daily case count began to climb once again.

Since the Public Health England announcement, countries worldwide have also increased mandatory testing before and/or after travel, as well as quarantine periods following travel, which can slow the spread of VOICs into regions where they have yet to establish. Enhanced measures to reduce transmission, including travel restrictions and lockdown measures that prohibit various activities, have subsequently reversed the rise in case numbers in many regions (Figure 4). Detection of immune-escape VOICs will pose specific challenges. Preliminary data suggest that vaccinated individuals who become infected are less likely to feel symptoms, making breakthrough infections less likely to be detected, although severe infections continue to occur. To detect and break transmission chains involving immune-escape variants, backwards contact tracing, as well as regular testing of asymptomatic individuals, may be essential tools, followed by enhanced control measures to prevent onward transmissions. Distinguishing escape mutations from chance breakthrough cases will also be a challenge. Flagging transmission chains that occur between vaccinated individuals, as well as whole-genome sequencing of breakthrough cases, will help detect new immune-escape VOICs earlier. Hampering this effort, however, is the fact that vaccination status, reason for sequencing, contact history, and case information (for example, severity) are typically not shared alongside genomic data, limiting the power to detect escape variants and prevent their global spread.

Finally, internationally coordinated genomic surveillance will be important for identifying which variants to include in future booster vaccines, with an eye to prioritizing immune-escape VOICs and highly divergent SARS-CoV-2 lineages.

The emergence of VOICs and the ongoing evolution of SARS-CoV-2 have potentially important implications for how best to control the pandemic. We first consider implications for detecting VOICs and then turn to different potential control measures that can help slow their emergence and spread.

**Detection and surveillance of VOICs**

Identifying VOICs as they emerge allows for the earlier implementation of control measures. Whole-genome sequencing followed by bioinformatic analysis can identify lineages that are increasing in frequency or that can infect previously immune individuals. Coupled with structural modelling of potential effects on viral function and the ability to evade antibodies, variants of interest (VOICs) can be identified (Box 1). Systematic surveillance and contact tracing are then needed to track a VOI to determine the mechanisms by which it is increasing in frequency, as well as whether it causes altered disease characteristics, such as increased transmission rates, longer periods of infectiousness, greater virulence, or the capacity to infect previously immune individuals. Assessing changes to the infectious period of a variant is, for example, critical for setting effective quarantine intervals.

Once a variant of interest is elevated to the status of VOC (Box 1), reducing its spread also requires effective genetic surveillance. Importantly, the higher the fraction of cases that are genetically typed, the easier it will be to contain a VOIC, because the variant is more likely to be detected early when there are fewer VOIC infections. For this reason, rapid PCR-based methods that detect a VOIC (for example the SGTF that allowed detection of B.1.1.7 in the UK) are an important tool, alongside whole genome sequencing, to detect VOICs early enough to isolate cases and prevent establishment.

**Control measures to slow the evolution of new variants**

Although we cannot generally reduce the rate of mutation, we may be able to reduce the risk that multiple mutations arise. In particular, we can provide a halo of protection around immunocompromised individuals by prioritizing vaccination of their households and caregivers, reducing their risk of exposure. If long-lasting infections do occur, enhanced surveillance and contact tracing should be considered to prevent onward

---

**Figure 3.** Weekly new cases per 100,000 people for Ontario, Canada. Data displayed as total count, those due to non-VOC, and those due to VOC (primarily B.1.1.7, as measured by a PCR test for the N501Y mutation). Plots are moving seven-day averages, using data from https://covid19.ontoisntable.ca/ontario-dashboard/. The decreasing total case count between 21 January, 2021 and early March, 2021 masked an underlying increase in the case count due to VOICs.

**Figure 4.** Weekly new cases per 100,000 people for Ontario, Canada. Data displayed as total count, those due to non-VOC, and those due to VOC (primarily B.1.1.7, as measured by a PCR test for the N501Y mutation). Plots are moving seven-day averages, using data from https://covid19.ontoisntable.ca/ontario-dashboard/. The decreasing total case count between 21 January, 2021 and early March, 2021 masked an underlying increase in the case count due to VOICs.
transmissions. Heightened measures to reduce onward transmission may also be warranted for lineages associated with higher-than-expected mutation rates (for example, mutants that alter proofreading capacity).

More broadly, we can reduce the rate of emergence of new VOCs and slow the spread of existing ones by reducing overall case numbers through vaccination at a global scale and by maintaining or enhancing the nonpharmaceutical interventions that have contributed to controlling the pandemic (case detection and isolation, contact tracing and quarantine, masking and personal distancing, and improved ventilation). Having low case numbers makes it easier to test and genotype a high fraction of cases and increases the efficacy of contact tracing measures to stop onward transmission. Furthermore, mathematical models predict that measures that reduce contact rates with susceptible individuals will not only slow the spread of SARS-CoV-2 overall but will also reduce the relative advantage of variants that have a transmission advantage. Thus, the measures taken to reduce contacts and limit the number of COVID-19 cases may have the added benefit of slowing the rate at which VOCs with a transmission advantage overtake the wild type. This predicted pattern, with selection weakening as stringency measures are increased, appears to be borne out in data for B.1.1.7 from England and British Columbia, Canada (Figure S2).

Importantly, the existence of VOCs means that scientists and public health officials cannot group all cases together when inferring the dynamics of SARS-CoV-2 and predicting future case loads. Each VOC will have its own reproductive number (or growth rate) and, consequently, different requisite measures for effective control. Additional restrictions (for example, further closures of schools, businesses, and leisure venues) can slow the spread and reduce the impact of VOCs, despite their higher transmission advantage. By flattening both non-VOC and VOC COVID-19 curves, more people can be vaccinated and protected before peak case numbers are reached.

Over the near future, epidemic control by nonpharmaceutical interventions will be replaced by control via vaccination, increasing selection in favour of immune escape. Reducing case numbers during this time will continue to be important. In particular, reducing exposure of immunized individuals to active cases should slow the evolution of immune-escape variants (Box 2) by reducing their selective advantage. Vaccinating whole communities or workplaces at once, coupling vaccination campaigns with temporary stay-at-home orders to reduce community transmission, and communicating the risks of exposure to vaccinated individuals are all potential actions that could reduce the evolution of immune escape variants, by reducing contact rates between infectious and immunized individuals. Having effective public health measures also reduces the likelihood that recombinant variants arise by driving down the chance of coinfection.

**Dosing strategies**

Data suggest that individuals who have received only a single dose of the Pfizer BNT162b2 vaccine or who have natural immunity elicit a weaker antibody response to VOCs than those who have received two vaccine doses, although T-cell responses were similar for variants and non-variants. Recent reports also indicate lower real-world effectiveness of vaccines against VOCs after one dose (Pfizer: 29.5% [95% CI: 22.9–35.5] for B.1.351 versus 16.9% [95% CI: 10.4–23.0] for B.1.1.7; AstraZeneca: 33.2% [95% CI: 8.3–51.4] for B.1.617.2 versus 49.2% [95% CI: 42.6–55.0] for B.1.1.7). These studies, though preliminary, also find that efficacy rises for all variants following the second dose. At this time, however, it is unknown how much of the rise in efficacy against VOCs after the second dose is due to the booster shot itself and how much is due to the maturation of the immune response given more time since the first dose. At face value, these data indicate that completing the recommended number of doses...
Figure 5. Selective advantage of B.1.1.7 declines with increasing social restrictions in the UK.
The same data as in Figure 2 but now the selective advantage of B.1.1.7, as measured by the weekly change in log(frequency SGTF/frequency of non-SGTF), is plotted against the stringency index of restrictions during that week in the UK (taken from http://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker). All data points corresponding to a frequency of less than 10% were excluded to ensure SGTF data predominantly reflect the presence of B.1.1.7.

for two-dose vaccines reduces the risk of transmission of partial-escape variants.

Over the short-term, however, the reduction in case numbers made possible by administering one dose to more people is likely to both save lives and reduce the chance that more resistance mutants arise in the first place65 (Box 2). As long as one dose provides sufficient immunity, reducing the burden of disease by vaccinating as many people as possible with one dose and delaying the second dose better prevents the evolution of immune escape variants than more slowly vaccinating the population with two doses at a short inter-dose interval16,67.

Another important factor to consider when determining the dosing regime comes from recent studies showing that delaying the second dose enhances immunity, with elevated antibody levels following a longer dosing interval (12 weeks) than a shorter interval (<6 weeks) for both the Pfizer68 and the AstraZeneca69 vaccines. Delayed second doses might thus increase the selective advantage of partial escape mutations over the short term but reduce it over the longer term given the more robust immune responses observed with a delayed booster68,69.

Reopening
As we move to reopen societies, rebuild economies, and resume social activities, contact rates will increase. This is expected to increase the selective advantage of VOCs that are more transmissible and/or more able to infect immunized individuals (Figure 5). Increased activity will also make contact tracing more difficult, as will the potential rise in asymptomatic cases among vaccinated individuals90. Coupling reopening with expanded use of rapid testing and environmental sampling would allow cases, and particularly asymptomatic breakthrough cases, to be identified and controlled. Border measures, with testing before, during, and after travel, along with quarantine measures, will continue to be important while cases remain high globally to slow the spread of immune-escape variants and other VOCs that emerge. Ultimately, though, variants will continue to emerge until cases are globally controlled through, for example, equitable access to vaccines. In the words of Ursula von der Leyen (President of the European Commission), “none of us will be safe until everyone is safe.”

Conclusions
As COVID-19 transitions from a pandemic to an endemic disease, VOCs present new global challenges to health by virtue of increased transmissibility and virulence and evasion of natural and vaccine-induced immunity. In this article we have explored the selective forces that shape how VOCs emerge and become established. We also identify possible steps that we can take to limit their emergence and, when they do arise, their impact. Moving forward, we must also consider how SARS-CoV-2 transmits to and amongst other animal species, placing both them and us at further risk. It will therefore be important to adopt a multidisciplinary One Health approach91 for future pandemic management that accounts for the interrelated nature of human, animal, and ecosystem health.

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.06.045.

ACKNOWLEDGEMENTS
The authors would like to thank Art Poon for providing the deviations from the molecular clock used in Figure S1B, Hongru Wang and Rasmus Nielsen for providing estimates of $\rho_t$ to estimate the mutation rate, and two referees for helpful suggestions. Funding was provided by the Natural Sciences and Engineering Research Council of Canada to S.P.O. (RGPIN-2016-03711) and T.D. (RGPIN-2013-217638), by Canada 150 Research Chair support for C.C., by the Public Health Agency of Canada to M.L., S.M., G.V.D., and N.H.O., by the CIHR 2019 Novel Coronavirus (COVID-19) rapid research program to J.W., and by the M.G. DeGroot Institute for Infectious Disease Research at McMaster University to J.D. and D.J.D.E.

DECLARATION OF INTERESTS
The authors declare no competing interests.

REFERENCES


