

On the evolutionary epidemiology of SARS-CoV-2

Troy Day¹, Sylvain Gandon², Sébastien Lion³, & Sarah P. Otto⁴

June 5, 2020

All authors contributed equally – authorship is alphabetical.

1. Department of Mathematics and Statistics, Department of Biology, Queen’s University, Kingston, Canada. day@queensu.ca
2. CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France. sylvain.gandon@cefe.cnrs.fr
3. CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France. sebastien.lion@cefe.cnrs.fr
4. Department of Zoology & Biodiversity Research Centre, University of British Columbia, Vancouver, Canada. otto@zoology.ubc.ca

Highlights

- The RNA virus SARS-CoV-2 is genetically variable, but there is no conclusive evidence that existing variants affect viral fitness or disease progression.
- Claims of positive selection must be tested against null models that account for the stochasticity of disease spread and founder events, which can mimic the action of selection.
- Modeling reveals how selection would act on SARS-CoV-2 mutations that alter viral transmission, disease progression, disease severity, or combinations of these traits.
- Direct evidence linking mutations to disease characteristics is needed before evolutionary ideas can guide public health interventions.
- If functional differences are verified, rapid typing would allow limited resources for mitigation measures to be tailored and targeted most appropriately.

eTOC

Evolutionary epidemiological models illustrate how selection might act on SARS-CoV-2. Considering the limited data, selection favors increased transmission, longer pre-symptomatic periods, fewer asymptomatic cases, and lower disease severity. Viral mutations are expected to affect combinations of these traits, however, making it challenging to predict the direction and disease impact of evolution.

There is no doubt that the novel coronavirus SARS-CoV-2 that causes COVID-19 is mutating and thus has the potential to adapt during the current pandemic. Whether this evolution will lead to changes in the transmission, the duration, or the severity of the disease is not clear. This has led to considerable scientific and media debate, from raising alarms about evolutionary change to dismissing it. Here we review what little is currently known about the evolution of SARS-CoV-2 and extend existing evolutionary theory to consider how selection might be acting upon the virus during the COVID-19 pandemic. While there is currently no definitive evidence that SARS-CoV-2 is undergoing further adaptation, continued, evidence-based, analysis of evolutionary change is important so that public health measures can be adjusted in response to substantive changes in the infectivity or severity of COVID-19.

Zoonotic pathogens, which have jumped from animal to human hosts, can result in enormous public health challenges because so little is known about the pathogen during the initial stages of an outbreak. The most important public health intervention for such pathogens is therefore to suppress transmission as much as possible. The current COVID-19 pandemic, caused by the SARS-CoV-2 coronavirus, provides a stark example. Wide-scale shifts in human social networks, from restrictions on travel to lockdowns of entire cities or countries, have been critical to the success of slowing the pandemic and reducing the number of deaths.

Because zoonotic pathogens are often poorly adapted following a host shift, it is also natural to ask how they will evolve in response to its novel human host and to medical and public health interventions. Examples where some evidence exists for adaptation following host shifts include myxoma virus in rabbits, and avian flu, Ebola, and Zika virus in humans [1]. With SARS-CoV-2, we might also expect further adaptation to its human host. For example, although SARS-CoV-2 is already able to bind the ACE2 receptors critical for entry into human cells, computational models and data have identified additional mutations that might further strengthen binding affinity [2].

In this essay, we explore the evolutionary potential for SARS-CoV-2, guided by available data and evolutionary models. At present, there is a lack of compelling evidence that any existing variants impact the progression, severity, or transmission of COVID-19 in an adaptive manner.

Models, however, indicate that natural selection can be strong and act on diverse aspects of SARS-CoV-2 as it spreads in its new human host. We argue for developing better strategies to detect, verify, and respond to evolutionary changes in COVID-19 that have important effects on human health and disease spread. Doing so will enhance the set of tools at our disposal for implementing effective public health measures.

Current empirical evidence

SARS-CoV-2 emergence

The growth of the human population has led to an increasing number of human-wildlife interactions, facilitating the movement of pathogens from animal hosts to humans (zoonoses) [3]. Viral spillover to a new species requires either pre-adaptation or rapid evolution of the proteins that dock and allow entry into new host cells. For SARS-CoV-2, six amino acids in the receptor-binding domain of the spike protein are critical for binding the host target receptor ACE2 and allowing infection in humans [4]. These critical spike protein residues are not all present in the most closely related coronavirus identified to date, RaTG13, sampled from the horseshoe bat, *Rhulophus affinis* (RaTG13 and SARS-CoV-2 are 96% similar at the nucleotide level), but they are found in coronavirus sampled from pangolin [4]. The SARS-CoV-2 genome shows no evidence of recent recombination, arguing against a recombinant origin involving pangolin [5, 6]. Given the poor sampling of coronaviruses from wildlife and the wide range of animals with similar ACE2 receptors (including pigs, ferrets, cats, and non-human primates [2]), it is likely that we simply have not identified the most closely related animal source [4], making it impossible to know what evolutionary changes happened immediately prior to or during the transition to humans.

Genetic Variation in SARS-CoV-2

Clues to the history of a disease can be obtained from its phylogenetic tree. Within humans, SARS-CoV-2 displays a star-like phylogeny with many long tip branches [7], as expected in a growing population. Based on genomic sampling over time, the substitution rate is estimated to be 0.00084 per site per year (www.nextstrain.org; [8]; 16 May 2020), 2- to 6-fold lower than the substitution rate for influenza (0.004-0.005 subst/site/yr for influenza A and 0.002 subst/site/yr

for influenza B [9]). Across its $\approx 30,000$ basepair genome (Nextstrain), SARS-CoV-2 thus undergoes roughly one genetic change every other week.

Genomic tracking tools like Nextstrain [8] allow us to analyze genetic variants very rapidly, as soon as their sequences become available during an outbreak [8]. Figure 1 illustrates nucleotide variation among the 5380 genomes available on May 16, 2020 (Figure 1A). Several of these are nonsynonymous and thereby alter the amino acid composition of viral proteins (Figure 1B).

Although Figure 1 reveals substantial genetic variation in SARS-CoV-2, it is unclear if these changes have any functional significance. Many are likely neutral or slightly deleterious to the virus [5, 10], having risen in frequency by chance when carried to new susceptible hosts. Mutations with no functional significance readily fluctuate in abundance, acting like genomic fingerprints that can be used to track viral geographic spread and to reconstruct epidemiological dynamics (see, e.g., May 8, 2020 Situation Report on international spread from Nextstrain). In the long run, deleterious mutations are expected to be eliminated, as seen in genetic comparisons between more distantly related coronavirus lineages [6]. However, even deleterious mutations can rise in abundance during an epidemic as long as their effective reproduction number remains larger than one.

More controversial is whether any of the nonsynonymous variants circulating among humans increases viral fitness. Box 1 describes four of the most prominent studies about adaptive SARS-CoV-2 mutations and the reasons for caution.

Natural Selection on SARS-CoV-2

Even though the adaptive significance of genetic variants remains to be established, we can use evolutionary theory to gain insights about how natural selection might act on disease characteristics. Modeling SARS-CoV-2 is currently challenging because we lack crucial information. For example, the fraction of cases that are asymptomatic, the relative infectiousness of asymptomatic individuals, and how these vary with age, are not yet well understood. We therefore first explore a general model, without specifying the exact parameter values. We then illustrate the dynamics using parameter values consistent with available data.

An extensive body of theory has been developed to understand the short- and long-term

evolution of pathogens [11–15]. There are two types of pathogen traits whose evolution is usually distinguished: antigenic traits and disease life-history traits [15].

Antigenic evolution refers to the mutational appearance and spread of viral genotypes that can escape existing immunity in the population [16]. For example, continued antigenic evolution is the reason why seasonal influenza vaccines must be periodically updated – the influenza virus eventually evolves so much that it escapes the immune response induced by the vaccine [17]. In the case of SARS-CoV-2, where no natural immunity previously existed [18] and no vaccine is yet in widespread use, natural selection for antigenic escape mutations will be very weak. Modeling protective immunity and escape mutations will, however, be an important step for future theory if SARS-CoV-2 becomes endemic to the human population.

Disease life-history evolution refers to the mutational appearance and spread of genotypes that cause different disease characteristics [19]. For example, the transmission rate of the virus, the length of the asymptomatic period that it causes, and the mortality induced by the infection are all disease life-history traits. As will be shown, these traits can be under strong selection, even for emergent diseases that are spreading rapidly in immunologically naive populations.

Modeling epidemiology

To make predictions about the evolution of pathogens, we must consider the potentially complex interaction between epidemiological and evolutionary dynamics. An important starting point is therefore to develop an appropriate model for the epidemiology of SARS-CoV-2.

Data from multiple countries suggest that the median duration of time between infection and the onset of symptoms (i.e., the incubation period) is approximately 5 days [20]. Also, infectiousness has been inferred to start ~2.5 days before symptom onset, with high transmission prior to the onset of symptoms [21]. Viral loads then drop after symptom onset, halving within ~2-4 days [21]. Together these results suggest that there is both an exposed but non-infectious stage and a pre-symptomatic stage that is highly infectious. Once infected, individuals either recover with relatively little medical intervention, or they progress to more severe disease and suffer a higher mortality rate. The overall case fatality for symptomatic individuals is estimated to be 1-2% [22], and the mean length of time from symptoms until death is approximately 18 days [23]. Finally, some infected individuals remain asymptomatic throughout the course of infection. Based on a

systematic review, Buitrago-Garcia *et al.* estimate that 29% of cases remain asymptomatic (CI: 23-37%) but note that this may be an overestimate due to publication biases and the requirement for at least one asymptomatic individual in many of the studies [24]. Regarding the source of new infections, their review suggests that pre-symptomatic cases account for about 40-60% of new infections, with <10% from asymptomatic individuals and the remainder from individuals who have developed symptoms [24]. As they highlight, these numbers remain highly uncertain. Figure 2 captures these key qualitative features of COVID-19, as we understand it, and can be used to examine the epidemiology and evolution of SARS-CoV-2 across a range of plausible parameters (see Supplemental Discussion for details).

We illustrate the epidemiological dynamics within a population in Figure 3 (dashed curves), using example parameter values. Without public health interventions (top panels) the rapid spread of SARS-CoV-2 generates a high prevalence of infections (red curves) followed by herd immunity, halting the local epidemic ('X'). In the bottom panels, we include periods of social distancing (grey regions), during which all transmission rates are reduced by 60%. Social distancing "flattens the curve", prolonging the duration of the epidemic and reducing peak health care demand. Even without accounting for the excess mortality that would result if hospitals became overwhelmed, social distancing also reduces the total number of deaths (compare inset bar heights in top and bottom panels). The reason is subtle. By reducing the height of the epidemic, public health measures reduce the number of infected individuals at the point in time when there no longer remain enough susceptible individuals to sustain the disease, even without social distancing, thereby lowering the number of subsequent infections and deaths. In addition, public health measures buy time for effective treatments and vaccines to be developed, which would lower mortality even further than shown in the bottom panels.

Modeling evolution

We next use the model in Figure 2 to explore how SARS-CoV-2 might evolve by considering the fate of mutations that alter the disease life history traits (e.g., transmission, disease progression, and/or virulence). Many functionally relevant mutations will be lost through stochasticity when they first arise, even if they are selectively advantageous. Indeed, the probability of a mutation escaping stochastic loss during the initial stages of an outbreak is $1 - 1/R_0$, where R_0 is the

number of new infections caused by an infected individual in a fully susceptible population (the initial ‘reproduction number’). Assuming the mutant R_0 is similar to that of the wildtype R_0 (estimated to be ≈ 2.5 [25]), this gives a probability of escaping stochastic loss of approximately 60%.

In the Supplemental Discussion, we describe the evolutionary dynamics of genotypes that avoid stochastic extinction. Figure 2 summarizes the resulting selection on each life history trait (derived analytically without specifying the parameter values, many of which remain uncertain). Of course, viral mutations can alter multiple traits simultaneously through pleiotropy. For example, mutations that increase viral replication rate might affect transmission rate (the β ’s), the rate of disease progression (the κ ’s), the fraction of asymptomatic cases (f), and/or virulence (α). The summed effect of selection on all such traits governs the fate of the mutant, as described by equation (S6).

Some intuition for the general results derived in the Supplemental Discussion can be obtained by considering the special case where individuals with symptoms are immediately isolated and where asymptomatic individuals do not contribute to new infections (so that β_I and β_A are negligible). In this case (and assuming a very transient exposed class), evolutionary change is driven by selection on the pre-symptomatic class, as they are then the major source of new infections. We can then describe the change in frequency of a mutation using a single equation (see Supplemental Discussion):

$$\frac{dq_P}{dt} = \underbrace{q_P(1 - q_P)}_{\text{genetic variance}} \underbrace{\left[\overbrace{S(1 - f)\Delta\beta_P}^{\text{transmission}} - \overbrace{S\beta_P\Delta f}^{\text{asymptomatic fraction}} - \overbrace{\Delta\kappa_P}^{\text{disease progression}} \right]}_{\text{selection coefficient}} \quad (1)$$

Here, q_P is the fraction of pre-symptomatic infections that harbour the new mutation, $\Delta\beta_P$ is the difference in pre-symptomatic transmission rate between the new mutation and wildtype, Δf is the difference in the fraction of asymptomatic individuals, and $\Delta\kappa_P$ is the difference in the rate at which pre-symptomatic individuals develop symptoms. Given that the ‘genetic variance’ in (1) is positive, the mutation will increase in frequency as long as its ‘selection coefficient’ is positive; that is, if it transmits better, leads to fewer asymptomatic cases, remains infectious longer, or some combination of the three.

We next use these results to examine the nature of selection acting on the life history traits of SARS-CoV-2. We start by considering selection acting on each life history trait on its own and then explore how mutations that affect multiple traits will evolve. Throughout, we use the full model, with selection described by equation (S6) in the Supplemental Discussion, using the special case (1) only to help explain the results. The accompanying *Mathematica* package provides all code and explores a range of parameters to confirm the robustness of the illustrated results (available from DRYAD: <https://doi.org/10.5061/dryad.5hqbzkh3g>).

Transmission rates – Selection favors genotypes that have a higher transmission rate, when considered as a trait on its own (no pleiotropy), with the strength of selection for increased transmissibility being proportional to the density of susceptible hosts, S . This result holds generally (equation S6) and can be seen most clearly in equation (1). This means that selection for genotypes that have a higher transmission rate will be strongest in dense populations with a large number of immunologically naive individuals [11, 13, 14], a prediction that has been verified experimentally in other systems [26, 27].

Public health interventions, like social distancing, will typically reduce the rate of contact among individuals and so reduce the transmission rate of all genotypes. As a result, social distancing will decrease the $\Delta\beta$'s (e.g., by 60% in Figure 3E) and so will weaken selection for increased transmission rates. That said, because social distancing will maintain a large population of susceptibles, S , for longer, this epidemiological feedback will increase the average strength of selection over the outbreak. In addition, social distancing spreads infections out over a longer period of time (i.e., it flattens the curve). This, in turn, allows selection to act over a longer time period, resulting in a larger cumulative amount of evolution (Box 2). Overall, adaptive mutations therefore typically reach a higher final frequency with social distancing than without (compare black curves in Figure 3A,E), although the net effect depend on the details.

Evolution, in turn, can affect the epidemiology of the disease, both in terms of infection prevalence (compare cumulative effects on the remaining number of susceptibles in Figure 3E with [solid blue] and without [dashed blue] evolution) and the cumulative number of deaths (inset bars in Figure 3E with [red] and without [pink] evolution). The public health importance of any evolutionary change depends also on the mutation's effect size (illustrated is a mutation that

increases transmission by 20%).

Finally, some mutations may become more strongly selected in the presence of social distancing if they allow for viral transmission despite the intervention (e.g., mutations affecting aerosolization or persistence in the environment). Other interventions, like rapid contact tracing and testing, essentially eliminate some transmission chains, but we would not expect this to affect the strength of selection among untraced cases, as long as genotypes are equally likely to be traced.

Asymptomatic infections – How selection acts on the proportion of asymptomatic cases, f , depends on the source of most new infections. Based on current estimates [24], <10% of new infections originate from asymptomatic individuals. As long as asymptomatic individuals play a smaller role in transmission than individuals who will develop symptoms, selection will favor a reduction in number of asymptomatic cases (lower f ; Figure 3B,F). In technical terms, as long as the ‘reproductive value’ of infections proceeding through the asymptomatic route in Figure 2 is lower than that for individuals who will eventually develop symptoms (see Supplemental Discussion), selection acts to lower f , because doing so increases the overall transmission of the virus. This result can be seen most clearly in expression (1) but holds in the full model as well (equation (S6)).

Equation (1) also reveals that the strength of selection on the fraction of asymptomatic cases is proportional to the density of susceptible hosts, just as with selection on transmission. As a result, selection for genotypes that produce a lower fraction of asymptomatic infections will also be strongest in dense populations with a large number of immunologically naive individuals.

Public health measures aimed at isolating people once they develop symptoms will weaken this selection, by reducing the reproductive value of infections leading to symptoms. The direction of selection will remain the same, however, as long as more infections arise from pre-symptomatic individuals than asymptomatic individuals.

That said, the conclusions of this subsection are those that are most subject to change in light of new data, because the reproductive values of the different infection routes strongly depend on how much time each class is infectious and on the relative transmission rates from asymptomatic, pre-symptomatic, and symptomatic cases, for which we lack direct estimates (see

Supplementary Discussion). Also uncertain is whether genetic mutations are available to the virus that could alter the fraction of asymptomatic individuals, which may depend more on characteristics of the host at the time of infection (age, health status, etc.).

Disease progression – Natural selection generally favours a lengthening of the pre-symptomatic phase (i.e., smaller values of κ_P), including mutations that reduce the morbidity of the disease so that people remain less aware of their symptoms. Simply put, the longer the pre-symptomatic period, the longer the duration over which the virus can transmit (Figure 3C,G).

While the special-case model without any disease class structure (equation (1)) suggests that selection on the length of the pre-symptomatic class should be insensitive to the number of susceptible individuals and to the extent of social distancing (which do not affect the length of the pre-symptomatic phase, $\Delta\kappa_P$), the full model is sensitive to both (for example, see social distancing effect in grey regions, Figure 3G) because of changes to the reproductive value of the pre-symptomatic class relative to the other types of infected individuals. In addition, by flattening the curve, public health measures allow selection to act over a longer period of time (Box 2). Thus, mutations that increase the length of the pre-symptomatic phase reach a higher frequency under social distancing (compare black curves in Figures 3C,G).

Disease virulence – In general, disease-induced mortality, α , is selected to decrease (Figure 3D,H; Supplemental Discussion), when considered as a trait on its own (no pleiotropy). However, this selection tends to be weak because most individuals with COVID-19 recover and the costs of mortality are paid only late in the course of the disease, after substantial transmission has already occurred (in the pre-symptomatic and early symptomatic phases). This can be seen most clearly in equation (1) where we have assumed that only pre-symptomatic individuals transmit. In that case, selection on virulence is entirely absent.

Again, public health measures change how natural selection acts upon SARS-CoV-2. In particular, the widespread recommendation to self-isolate once symptoms appear has likely weakened direct selection against virulence, although this effect must be small given that direct selection against virulence was already weak (as seen in Figure 3D,H). Moreover, the benefits of self-isolation in terms of disease containment massively outweigh the small potential costs of reduced selection against virulence (compare inset bar charts in Figure 3D,H).

Pleiotropic mutations – Viral mutations that affect the course of the disease will often impact more than one life-history trait. The fate of such pleiotropic mutations will depend on the sum of the selective effects on each trait (S6). Furthermore, accurate predictions about the evolutionary future of SARS-CoV-2 require data about the impact of mutations on virus dynamics and viral load and disease progression within an individual and how these relate to the life history traits governing the spread of the pathogen. In the absence of such data, modeling can help guide our understanding about the range of possibilities.

A crucial question is how virulence will evolve [28]. As discussed above, direct selection on virulence is weak (see Figure 3D,H). Thus, virulence evolution will be driven largely by the indirect effects of pleiotropy. In Figure 4, we consider two potential examples. First, consider mutations that couple a higher transmission rate, the β 's, with higher mortality, α (positive pleiotropy, Figure 4A,C), as might occur if mutations increase viral replication rates. In this case, evolution will lead to higher mortality (see inset bars), as an indirect consequence of selection for increased transmission (see Supplemental Discussion and also [12, 29]). Alternatively, consider a mutation that alters tissue tropism such that the disease tends to preferentially infect cells of the upper respiratory tract, rather than the lower respiratory tract. Such infections could lead to a higher transmission rate but be less virulent (negative pleiotropy) [30]. This would generate indirect selection for lower mortality rates (Figure 4B,D).

Of course, pleiotropy may link virulence evolution with any of the life history traits of SARS-CoV-2. For example, selection to prolong the infectious stage, the κ 's, might reduce virulence as a side consequence (e.g., if a weaker immune reaction is elicited). Or selection for more symptomatic cases (lower f) may lead to a pleiotropic increase in virulence. Pleiotropy is hard to predict, which is why it is not possible to say with any confidence whether evolution of COVID-19 will translate into meaningful effects for patients (positive or negative) or for the spread of the disease.

Conclusion

In the midst of the COVID-19 pandemic, there is great uncertainty about the course of the disease, and this uncertainty is heightened by concerns about how the virus might evolve. It would be surprising if no mutations were available that could increase fitness of SARS-CoV-2 in

its novel human host. But considering the array of life history traits under selection and the myriad ways that mutations can link these traits (Figure 2), it is difficult to make clear predictions for the impact on disease. Before we incorporate evolutionary considerations into mitigation efforts, given the costs of mistakes – both in human lives and in public health dollars – we must hold claims that a variant increases viral fitness to a high evidentiary standard. This requires both that neutral explanations be rejected and that a clear link be documented between the mutation and characteristics of the disease. To date, none of the SARS-CoV-2 variants have met this standard (Box 1).

Of course, the absence of evidence is not the same as evidence of absence. At present, we know very little about the scope for adaptation or the functional significance of existing genetic variants. This knowledge gap is exacerbated by the lack of accessible data linking disease outcomes with genetic variants. The fact that the virus already displays effective human-to-human transmission might mean there is little opportunity for further adaptation. Likewise, the low mutation rate of SARS-CoV-2 compared to influenza suggests that it may evolve more slowly in response to selection, although the relative rates of evolution depend too on how many sites are targeted by selection (SARS-CoV-2's genome is double in size). On the other hand, we also know that the virus often fails to infect close contacts [31], suggesting that there is ample scope to increase transmission further. Moreover, the large number of viruses circulating within a patient implies that every possible genomic mutation is likely to arise over the course of an infection (median estimate of 100,000 viral copies per mL of saliva [32]). Thus, adaptation in response to strong selection for survival and transmission in human hosts could occur very rapidly and may have done so when it first switched into humans.

This paper is intended to provide a framework for thinking about the potential evolutionary routes that SARS-CoV-2 might take and to dispel some of the current misinformation that is circulating in the media. For example, in the absence of pleiotropy, the mortality rate due to COVID-19 is actually either selectively neutral or selected against. Furthermore, selection for increased transmission rate is at its strongest during the initial phase of an epidemic, when spread is exponential, even though one might have initially thought that selection would be stronger later, when host immunity is more widespread. Our analysis also shows that selection likely favors viruses that progress slowly towards disease and whose symptoms remain mild for longer,

because such viruses will be transmitted more before people are aware that they are infected.

This contrasts strongly with Ebola, where a major transmission route was contact with the dead, which can strongly select for greater virulence [33, 34]. On the other hand, if transmission is pleiotropically coupled with faster rates of viral replication, and the latter leads to greater virulence, then selection for more efficient transmission could result in an evolutionary increase in virulence (as illustrated in Figures 4A,C).

As with any theoretical analysis, our predictions rely on some biologically informed simplifications. For example, we have neglected within-host evolution and the roles of age- and spatial structure in the transmission process. Likewise, we have largely ignored the influence of stochasticity, despite the fact that chance events like founder effects, genetic drift, and super-spreading can be particularly important in emerging diseases. It is also important to emphasize that much remains unknown about SARS-CoV-2. As our understanding of the virus improves and new data emerge, it will be possible to refine predictions and explore other scenarios for the short- and longer-term evolution of viral traits. Experimental data using cell lines and animal models will help reveal the pleiotropic effects of mutations in SARS-CoV-2. Phylogenetic methods will help reveal which, if any, genetic changes have been driven by positive selection. Such methods include classic approaches searching for sites that undergo amino acid changes more often than expected (e.g., [35]), as well as newer methods that use the shape of the phylogeny to infer the effect of genetic variants on transmission rates and virulence [36]. A key challenge for future studies is to determine whether inferences are robust to the extreme stochasticity we are seeing with COVID-19, where travelers seed infections on different dates from different sources, where a carrier can pass the virus on to dozens of individuals or to none, and where human behavioral responses vary widely in space and time.

A more powerful method to detect the functional importance of mutations would be to leverage the enormous world-wide effort to sequence SARS-CoV-2 genomes. We recommend that metadata be sought when viral samples are collected that provide information about disease outcomes (time since symptom onset, severity of symptoms at time of sampling) and, where available, with data on hospitalization rates, death rates, and transmission (if contact tracing is conducted). Relating genomic variants with patient health and epidemiological data is the most

direct way to establish whether mutations are functionally important or not (as recently investigated by [37] using hospitalization data). This data is also needed to apply many of the new phylogenomic methods, as argued by [1].

Finally, if functionally important mutations are verified, how might this then inform intervention strategies? While the answer to this question will depend on the details, it can be helpful to consider a few possibilities to reinforce the value of tracking evolution during a pandemic. If a variant is found to be more transmissible, then its spread could be mitigated by directing limited public health resources to populations in which this variant is most strongly selected (i.e., dense, immunologically naive, populations). Likewise, by using PCR tests to distinguish functionally important variants, public health officials could know which variants are circulating in their communities, which would allow them to loosen or tighten social interventions where milder or more severe forms arise. By knowing which variant a patient carries, doctors could also then adjust drug regimens, applying them earlier and/or more aggressively in higher risk situations (i.e., personalized medicine). Even intentional inoculations are possible if a particularly mild variant is discovered. But first we need to know which, if any, variants are functionally important to COVID-19, with direct evidence linked to health outcomes.

Acknowledgements

We are grateful to four anonymous reviewers, Ailene MacPherson, Leithen M'Gonigle, Arne Mooers, Berni Crespi, and Michael Whitlock for helpful comments on previous drafts that helped refine this work. Funding was provided by the Natural Sciences and Engineering Research Council of Canada to TD (RGPIN-2013-217636) and to SPO (RGPIN-2016-03711) and by the Agence Nationale de la Recherche to SG (ANR-17-CE35-0012) and SL (ANR-16-CE35-0012-01).

Declaration of Interests: The authors declare no competing interests.

Box 1: Searching for adaptive mutations in SARS-CoV-2

Within weeks after the initial reports emerging from Wuhan, China, of a new respiratory illness in December 2019, scientists had already started searching for signatures of adaptation to humans within the genomes of SARS-CoV-2. One of the earliest studies to appear identified two strains of SARS-CoV-2 circulating in Wuhan (strains 'L' and 'S') and suggested that they had functional consequences, with the former being more 'aggressive' [5]. This claim was widely picked up in the media and led to considerable speculation that evolutionary change could result in COVID-19 becoming more severe. However, this inference was based solely on the frequencies of the two variants, and the stochastic occurrence of mutations on the basal branches of a star-like phylogeny for a spreading disease can fully account for the frequency data [10]. The original authors have since acknowledged this more parsimonious explanation [5].

Similarly, a recent study [37] found that the non-synonymous mutation D614G in the spike gene (see peak in "S" in Figure 1B) has been increasing in frequency in multiple countries. Although this pattern is expected if the mutation were selectively advantageous, it can also be explained by purely neutral sampling processes. Even without selection, a parallel increase in the frequency of a mutation across multiple countries is expected if new disease outbreaks are first seeded by travelers from a geographic location with a low mutant frequency (e.g., China) followed by travelers from a location where the mutation is (by chance) at a high frequency (e.g., Italy). Definitive evidence of selection must therefore reject such a neutral explanation, which was not done [37]. On the other hand, [37] did look for corroborating evidence of selection in hospitalization rates but did not find a difference associated with D614G. The authors did find a difference in viral load, but this finding is hard to interpret given that days since symptom onset (a major determinant of viral load [21]) was not controlled and may have changed over time as testing became more available. Additional data are thus needed before concluding that selection is acting upon D614G.

Another genetic change that has received media attention involves a 382 nucleotide deletion in ORF8 found in multiple COVID-19 patients in Singapore. Although no direct evidence exists that this deletion was positively selected, similar deletions have been found in other coronaviruses, including SARS variants that arose during the 2003-2004 outbreak [38]. Cell culturing demonstrated that one of these earlier SARS deletions reduced the rate of viral replication [39]. The repeated appearance of such deletions is intriguing, but direct evidence is needed linking such deletions to disease outcomes and/or transmission rates.

Finally, [40] used sister clade comparisons to examine whether any of 31 specific mutations identified in the SARS-CoV-2 genome, including the D614G mutation mentioned above, are associated with an increased transmission rate. Their results show that most mutations are found in clades that, if anything, are associated with reduced transmission, concluding that there is no evidence for positively selected alleles. However, sister clade comparisons lack power and are biased against finding derived characters that boost the growth of a lineage.

Definitive conclusions await further monitoring and testing, accounting for the null expectations with a rapidly expanding zoonotic disease.

Box 2: Evolution and flattening the curve

The key public health interventions used against COVID-19 focus on reducing contact rates, isolating travelers and sick individuals, and contact tracing. These interventions aim to reduce transmission (the β 's) and 'flatten' the trajectory describing the time-course of infections or, even better, squash it to zero. These interventions are essential for health care systems to manage the influx of cases and to save lives.

Flattening the epidemic curve has a secondary effect though: by prolonging the duration of the epidemic, it increases the time period over which evolutionary change accumulates. As seen in Figures 3 and 4 (compare top and bottom panels), these interventions often result in a greater total change in gene frequency. This finding is perhaps not as obvious as it first seems. Evolutionary adaptation requires that new infections be generated and old ones lost at rates that differ among genotypes. Thus, it seems reasonable to expect that the cumulative amount of evolutionary change in gene frequency that occurs over an epidemic is positively related to the total number of new infections that occur (i.e., the outbreak size). And since public health interventions typically reduce outbreak size we might expect them to also reduce the total amount of evolution. However, this is not true. The rate of evolution is determined by the *difference* in growth rate of infections carrying the two different alleles (this is the selection coefficient in equation (1)) whereas the total outbreak size is determined by the values of these growth rates themselves. This is what evolutionary biologists refer to as the distinction between relative versus absolute fitness. So interventions like social distancing that flatten the curve can slow the rate at which an epidemic grows without slowing the rate at which evolution occurs. Because such interventions typically extend the duration of the epidemic, all else equal, they will thereby result in a greater total amount of evolutionary change. Of course, all else need not be equal, because social distancing can also affect the strength of selection (see main text). The net outcome will therefore depend on the balance of these effects.

In addition, reducing the total number of infections will reduce the input of SARS-CoV-2 mutations, which can slow adaptation, especially if complex mutations underlie fitness gains. Furthermore, even if rapid epidemics end the potential for evolution locally (see 'X's in Figures 3 and 4), evolution will continue globally as long as COVID-19 is still circulating in the human population. To determine the net impact of public health measures on evolutionary change in this context requires models that consider the appearance of mutations and their geographic spread.

FIGURE LEGENDS

Figure 1: Variability among SARS-CoV-2 genomes. Genetic diversity segregating among SARS-CoV-2 genomes (from Nextstrain; [8]). Horizontal axis is genomic location and vertical axis is entropy, an information-based measure that highlights sites exhibiting the most genetic variation: A at the nucleotide level, B at the amino acid level.

Figure 2: Epidemiological model for COVID-19. Susceptible individuals, S , enter the exposed class, E , upon infection after contact with infected individuals (as indicated by dashed curves, with transmission rates β_i). A proportion of new infections, f , remain asymptomatic, A , while the remainder become pre-symptomatic, P . The latter eventually progress to the symptomatic stage, I , which die from the disease at rate α (referred to as ‘virulence’). All other individuals eventually recover, R , and are assumed to be immune. Transition rates between disease classes are denoted by κ_i . As shown in the Supplemental Discussion, selection acting on these traits favors increased transmission and a briefer interval between exposure and infectiousness (red parameters are selected to increase). Selection also favors mutations that keep individuals in the infectious stage longer (green parameters are selected to decrease), including reduced virulence. As long as pre-symptomatic and symptomatic individuals are the major source of new infections, selection also favors a reduction in the proportion of asymptomatic individuals (f).

Figure 3: Simulations of COVID-19 evolution without pleiotropy. Evolutionary and epidemiological dynamics of a mutation that only affects a single trait, either in the absence (top panels) or presence (bottom) of periodic social distancing. Parameters of the wild type virus (r) are chosen to be roughly consistent with available data: $\beta_P^r = 3\beta_I^r = 10\beta_A^r = 1$, $\kappa_E^r = 0.25$, $\kappa_P^r = 1$, $\kappa_I^r = 0.2$, $\kappa_A^r = 0.11$, $f^r = 0.2$, and $\alpha^r = 0.005$. This yields a basic reproduction number $R_0^r \approx 2.3$. A mutant allele increases transmission in panels A and E (all transmission rates multiplied by 1.2), decreases the fraction of asymptomatic cases in panels B and F ($f^m = 0.1$), progresses more slowly through the pre-symptomatic stage in panels C and G ($\kappa_P^m = 0.67$), and decreases virulence in panels D and H ($\alpha^m = 0$). In the latter case, mutants that reduce mortality do spread, but selection is very weak and the effects are hardly visible. Grey regions indicate periods of effective social distancing (all transmission rates are reduced by 60%). Curves show the numbers of infected (red) and susceptible individuals (blue), measured as a fraction of the initial number of susceptibles, as well as the frequency of the mutation (black). Solid curves are

with evolution (dashed without for reference). Inset bar chart shows cumulative deaths, with ticks at 1% intervals (pink: without evolution; red: with evolution).

Figure 4: Simulations of COVID-19 evolution with pleiotropy. Evolutionary and epidemiological dynamics of a mutation with pleiotropic effects on both transmission (all transmission rates multiplied by 1.2) and virulence, either doubling virulence ($\alpha^m = 0.01$, panels A,C) or eliminating it ($\alpha^m = 0$, panels B,D). See Figure 3 for additional details.

References

1. Geoghegan, J.L. and Holmes, E.C. (2018). The phylogenomics of evolving virus virulence. *Nat. Rev. Gen.* *19*, 756–769.
2. Wan, Y., Shang, J., Graham, R., Baric, R.S., and Li, F. (2020). Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* *94*, e00127–20.
3. Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature* *451*, 990–993.
4. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* *26*, 450–452.
5. Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., et al. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review* *nwaa036*.
6. Boni, M.F., Lemey, P., Jiang, X., Lam, T.T.Y., Perry, B., Castoe, T., Rambaut, A., and Robertson, D.L. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *bioRxiv*
<https://www.biorxiv.org/content/early/2020/03/31/2020.03.30.015008>.
7. Mavian, C., Marini, S., Manes, C., Capua, I., Prosperi, M., and Salemi, M. (2020). Regaining perspective on SARS-CoV-2 molecular tracing and its implications. *medRxiv*
<https://www.medrxiv.org/content/10.1101/2020.03.16.20034470v1>.
8. Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., and Neher, R.A. (2018). Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* *34*, 4121–4123.
9. Bedford, T., Riley, S., Barr, I.G., Broor, S., Chadha, M., Cox, N.J., Daniels, R.S., Gunasekaran, C.P., Hurt, A.C., Kelso, A., et al. (2015). Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* *523*, 217–220.
10. Maclean, O.A., Orton, R., Singer, J.B., and Robertson, D.L. (2020). Response to: On the

origin and continuing evolution of SARS-CoV-2. <http://virological.org/t/response-to-on-the-origin-and-continuing-evolution-of-sars-cov-2/418>.

11. Lenski, R.E. and May, R.M. (1994). The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. *J. Theor. Biol.* *169*, 253–265.
12. Frank, S.A. (1996). Models of Parasite Virulence. *Q. Rev. Biol.* *71*, 37–78.
13. Day, T. and Proulx, S.R. (2004). A general theory for the evolutionary dynamics of virulence. *Am. Nat.* *163*, E40–E63.
14. Day, T. and Gandon, S. (2007). Applying population-genetic models in theoretical evolutionary epidemiology. *Ecol. Lett.* *10*, 876–888.
15. Gandon, S., Day, T., Metcalf, C.J.E., and Grenfell, B.T. (2016). Forecasting epidemiological and evolutionary dynamics of infectious diseases. *Trends Ecol. Evol.* *31*, 776–788.
16. Smith, D.J., Lapedes, A.S., de Jong, J.C., Bestebroer, T.M., Rimmelzwaan, G.F., Osterhaus, A.D., and Fouchier, R.A. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science* *305*, 371–376.
17. Yamayoshi, S. and Kawaoka, Y. (2019). Current and future influenza vaccines. *Nat. Med.* *25*, 212–220.
18. Amanat, F., Stadlbauer, D., Strohmeier, S., Nguyen, T.H., Chromikova, V., McMahon, M., Jiang, K., Arunkumar, G.A., Jurczynszak, D., Polanco, J., et al. (2020). A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0913-5>.
19. Dieckmann, U., Metz, J.A., and Sabelis, M.W. (2005). Adaptive dynamics of infectious diseases: in pursuit of virulence management, volume 2 (Cambridge University Press).
20. Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., and Lessler, J. (2020). The incubation period of Coronavirus Disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* *172*, 577–582.
21. He, X., Lau, E.H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y.,

- Tan, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* *26*, 672–675.
22. Wu, J.T., Leung, K., Bushman, M., Kishore, N., Niehus, R., de Salazar, P.M., Cowling, B.J., Lipsitch, M., and Leung, G.M. (2020). Estimating clinical severity of COVID-19 from transmission dynamics in Wuhan, China. *Nat. Med.* *26*, 506–510.
23. Verity, R., Okell, L.C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P.G.T., Fu, H., et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* *20*, 669–77.
24. Buitrago-Garcia, D.C., Egli-Gany, D., Counotte, M.J., Hossmann, S., Imeri, H., Salanti, G., and Low, N. (2020). The role of asymptomatic SARS-CoV-2 infections: rapid living systematic review and meta-analysis. medRxiv
<https://www.medrxiv.org/content/early/2020/04/29/2020.04.25.20079103>.
25. Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., and Eggo, R.M. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* *20*, 553–58.
26. Berngruber, T.W., Froissart, R., Choisy, M., and Gandon, S. (2013). Evolution of virulence in emerging epidemics. *PLoS Pathog.* *9*, 1–8.
27. Berngruber, T.W., Lion, S., and Gandon, S. (2015). Spatial structure, transmission modes and the evolution of viral exploitation strategies. *PLoS Pathog.* *11*, 1–13.
28. Cressler, C.E., McLeod, D.V., Rozins, C., Van Den Hoogen, J., and Day, T. (2016). The adaptive evolution of virulence: a review of theoretical predictions and empirical tests. *Parasitology* *143*, 915–930.
29. Day, T. (2003). Virulence evolution and the timing of disease life-history events. *Trends Ecol. Evol.* *18*, 113–118.
30. Tellier, R. (2006). Review of aerosol transmission of influenza A virus. *Emerg. Infect. Dis.* *12*, 1657–62.
31. Luo, L., Liu, D., Liao, X.l., Wu, X.b., Jing, Q.l., Zheng, J.z., Liu, F.h., Yang, S.g., Bi, B., Li,

- Z.h., et al. (2020). Modes of contact and risk of transmission in COVID-19 among close contacts. medRxiv <https://www.medrxiv.org/content/10.1101/2020.03.24.20042606v1>.
32. To, K.K.W., Tsang, O.T.Y., Leung, W.S., Tam, A.R., Wu, T.C., Lung, D.C., Yip, C.C.Y., Cai, J.P., Chan, J.M.C., Chik, T.S.H., et al. (2020). Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect. Dis.* *20*, 565–74.
33. Day, T. (2002). Virulence evolution via host exploitation and toxin production in spore-producing pathogens. *Ecol. Lett.* *5*, 471–476.
34. Sofonea, M.T., Aldakak, L., Boullosa, L.V., and Alizon, S. (2018). Can Ebola virus evolve to be less virulent in humans? *J. Evol. Biol.* *31*, 382–392.
35. Nielsen, R. and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* *148*, 929–936.
36. Rasmussen, D.A. and Stadler, T. (2019). Coupling adaptive molecular evolution to phylodynamics using fitness-dependent birth-death models. *eLife* *8*, e45562.
37. Korber, B., Fischer, W., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Foley, B., Giorgi, E., Bhattacharya, T., Parker, M., et al. (2020). Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. bioRxiv <https://www.biorxiv.org/content/early/2020/05/05/2020.04.29.069054>.
38. Chinese SARS Molecular Epidemiology Consortium (2004). Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* *303*, 1666–1669.
39. Muth, D., Corman, V.M., Roth, H., Binger, T., Dijkman, R., Gottula, L.T., Gloza-Rausch, F., Balboni, A., Battilani, M., Rihtarič, D., et al. (2018). Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci. Rep.* *8*, 1–11.
40. Van Dorp, L., Richard, D., Tan, C.S.C, Shaw, L.P., Acman, M. and Balloux, F. (2020). No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. bioRxiv <https://www.biorxiv.org/content/10.1101/2020.05.21.108506v1.full.pdf>

Supplementary Information

To understand the evolution of SARS-CoV-2, we describe changes over time in the number of individuals for each disease class (S , E , A , P , I , R) and genotype (*) using the following set of differential equations (see Figure 2 for a description of the notation and reference [S1] for an introduction to modeling disease dynamics:

$$\frac{dS}{dt} = -S \sum_* (\beta_P^* P^* + \beta_I^* I^* + \beta_A^* A^*)$$

$$\frac{dE^*}{dt} = S(\beta_P^* P^* + \beta_I^* I^* + \beta_A^* A^*) - \kappa_E^* E^*$$

$$\frac{dA^*}{dt} = f^* \kappa_E^* E^* - \kappa_A^* A^*$$

(S1)

$$\frac{dP^*}{dt} = (1 - f^*) \kappa_E^* E^* - \kappa_P^* P^*$$

$$\frac{dI^*}{dt} = \kappa_P^* P^* - (\alpha^* + \kappa_I^*) I^*$$

$$\frac{dR}{dt} = \sum_* (\kappa_I^* I^* + \kappa_A^* A^*)$$

The last equation assumes that previously infected individuals are resistant to further infection, which is expected but not currently known. This epidemiological system yields the following expression for the basic reproduction number of a genotype (*):

$$R_0^* = f^* \frac{\beta_A^*}{\kappa_A^*} + (1 - f^*) \left(\frac{\beta_P^*}{\kappa_P^*} + \frac{\beta_I^*}{\alpha^* + \kappa_I^*} \right)$$

(S2)

Equations (S1) allow for an arbitrary number of genotypes, but for simplicity we focus on two: the resident (where $* = r$) and a mutant ($* = m$). Within each class X of infected individual (e.g., pre-symptomatic, $X = P$), the frequency of the mutation, q_X , evolves according to:

$$\frac{dq_X}{dt} = \frac{d(X^m / (X^m + X^r))}{dt} \quad (\text{S3})$$

For example, consider a mutation that changes all of the transmission rates by $\Delta\beta$. Calculating the change in frequency of the mutant within each disease class gives:

$$\begin{aligned} \frac{dq_E}{dt} = & S \left(\frac{I}{E} \beta_I (q_I - q_E) + \frac{A}{E} \beta_A (q_A - q_E) + \frac{P}{E} \beta_P (q_P - q_E) \right) \\ & + S \Delta\beta (1 - q_E) \left(\frac{A}{E} q_A + \frac{I}{E} q_I + \frac{P}{E} q_P \right) \\ \frac{dq_A}{dt} = & f \kappa_E (q_E - q_A) \frac{E}{A} \end{aligned} \quad (\text{S4})$$

$$\frac{dq_P}{dt} = (1 - f) \kappa_E (q_E - q_P) \frac{E}{P}$$

$$\frac{dq_I}{dt} = \kappa_P (q_P - q_I) \frac{P}{I}$$

The second line of the first equation describes selection on the mutation, which favors variants with an increased transmission rate ($\Delta\beta > 0$). The remaining terms describe movement of resident and mutant infections among the different classes.

To understand the net effect of selection and movement, especially in more complicated cases where mutations affect multiple life-history traits, we must approximate (S1). One approach assumes weak selection and focuses on the early growth phase of the

epidemic, assuming that the number of susceptible individuals does not decline fast relative to the transition rates among disease classes. Then, treating S as a constant in equations (S1), the dynamics of viral genotype $*$ within infected individuals (E^* , A^* , P^* , I^*) become a linear system of equations that can be described in matrix form, \mathbf{M}^* , and analysed using standard approaches for stage-structured models (see, e.g., reference [S1]). The change in frequency of a mutation is then approximated by its effect on the leading eigenvalue (λ). For mutations that have a small effect on disease attributes (z), we can write:

$$\frac{d\lambda}{dz} = \vec{v}^T \frac{d\mathbf{M}}{dz} \vec{u} \quad (\text{S5})$$

where \vec{v} and \vec{u} are the left and right eigenvectors of \mathbf{M} (describing the reproductive values and the class frequencies, respectively, for the resident genotype, defined such that $\vec{v}^T \vec{u} = 1$), while $\frac{d\mathbf{M}}{dz}$ describes the change in \mathbf{M}^* caused by the mutation. Evaluating (S5) gives the effect of a mutation on the rate of spread of the disease:

$$\begin{aligned} \frac{d\lambda}{dz} = & S v_E (\Delta\beta_P u_P + \Delta\beta_I u_I + \Delta\beta_A u_A) - (\Delta\alpha + \Delta\kappa_I) u_I v_I - \Delta\kappa_A u_A v_A \\ & - \Delta f \kappa_E u_E \{v_P - v_A\} \\ & + \Delta\kappa_E u_E \{(v_P(1-f) + f v_A) - v_E\} - \Delta\kappa_P u_P \{v_P - v_I\} \end{aligned} \quad (\text{S6})$$

Equation (S6) tells us how selection acts to shape all of the life-history parameters, as illustrated in Figure 2. Because the eigenvector elements are non-negative, selection always favors higher transmission (positive $\Delta\beta_X$ terms), lower mortality (negative $\Delta\alpha$), and slower recovery (negative $\Delta\kappa_I$ and $\Delta\kappa_A$). As detailed in an accompanying *Mathematica* package (available for download from DRYAD: <https://doi.org/10.5061/dryad.5hqbzkh3g>), the terms in curly braces are also positive in a growing population under reasonable assumptions for SARS-CoV-2 (no infections

arise from recently exposed individuals, few arise from asymptomatic individuals, while pre-symptomatic individuals are highly infectious and expected to give rise to more total new cases than later-stage symptomatic individuals who have less time to pass on the disease) [S2]. Selection during the growth phase of the disease then favors fewer transitions to the asymptomatic class (negative Δf), a quicker progression out of the non-infectious exposed class (positive $\Delta \kappa_E$), and a slower progression out of the pre-symptomatic class (negative $\Delta \kappa_P$).

Note that the predictions on the evolution of the asymptomatic class f are sensitive to the reproductive values of asymptomatic individuals relative to pre-symptomatic ones. For example, in Figure 3, we assumed that transmission rates for asymptomatic individuals are only $1/10^{\text{th}}$ that of pre-symptomatic cases, but if transmission rates were equal then selection reverses and favors more asymptomatic cases (higher f , see additional figures in DRYAD package). On the other hand, we also assumed that transmission occurred over a longer period – 9 days, on average – for asymptomatic cases versus 6 days for individuals who will eventually show symptoms; if the period of asymptomatic transmission were shorter, selection would again favor a reduction in number of asymptomatic cases (lower f).

An alternative approximation is possible when the time spent in the exposed class is short enough that the E class can be ignored (κ_E high: newly infected individuals enter directly into the pre-symptomatic or asymptomatic classes), we can then revise (S1) by dropping $\frac{dE^*}{dt}$ and replacing:

$$\begin{aligned}\frac{dA^*}{dt} &= f^* S(\beta_P^* P^* + \beta_I^* I^* + \beta_A^* A^*) - \kappa_A^* A^* \\ \frac{dP^*}{dt} &= (1 - f^*) S(\beta_P^* P^* + \beta_I^* I^* + \beta_A^* A^*) - \kappa_P^* P^*\end{aligned}\tag{S7}$$

When self-isolation of symptomatic cases is effective and new infections emerge

primarily from the pre-symptomatic class ($\beta_P^* \gg \beta_I^*, \beta_A^*$), the pre-symptomatic class will predominate the seeding of new infections, and evolutionary changes in this class will be inherited by descendent infections. We can thus approximate the evolutionary trajectory by describing the mutant dynamics in this source class, P . Plugging dP^*/dt from (S7) into (S3) for the pre-symptomatic class ($X^* = P^*$) gives the evolutionary equation (1) discussed in the text (assuming weak selection and dropping lower order terms such as $\Delta\beta_P \Delta f$).

Numerical solutions – Figures 3 and 4 illustrate the disease dynamics given by equation (S1) for parameters chosen to be roughly consistent with the references cited in the main text. All individuals are assumed to be initially susceptible ($S = 1$, measured in units of the initial population size), into which a small fraction of exposed individuals is introduced ($E^* = 10^{-5}$). The introduced virus is assumed to be either genetically homogeneous (no evolution case; dashed curves) or to consist of 99% resident and 1% mutant genotypes (allowing evolution; solid curves). Although we consider a deterministic model, the pathogen population is doomed to go extinct locally because the resource (the density of susceptible hosts) is assumed to decrease over the time scale of the pandemic (we neglect human reproduction, migration, and waning immunity). We thus assume the pathogen goes locally extinct if the frequency of infected individuals falls below 10^{-7} (the ‘X’ in Figures 3 and 4). Simulation code for Figures 3 and 4, as well as additional figures illustrating different parameter choices, is available in the DRYAD package. Results are similar to those shown except where noted (see above on asymptomatic infections).

References

- S1. Otto, S.P. and Day, T. (2007). *A Biologist’s Guide to Mathematical Modeling in Ecology and Evolution* (Princeton, NJ: Princeton University Press).
- S2. Buitrago-Garcia, D.C., Egli-Gany, D., Counotte, M.J., Hossmann, S., Imeri, H.,

Salanti, G., and Low, N. (2020). The role of asymptomatic SARS-CoV-2 infections: rapid living systematic review and meta-analysis. medRxiv
<https://www.medrxiv.org/content/early/2020/04/29/2020.04.25.20079103>.