Lockdowns exert selection pressure on overdispersion of SARS-CoV-2 variants

Nielsen, Bjarke Frost; Eilersen, Andreas; Simonsen, Lone; Sneppen, Kim

Published in:
Epidemics

DOI:
10.1016/j.epidem.2022.100613

Publication date:
2022

Document version
Publisher’s PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Lockdowns exert selection pressure on overdispersion of SARS-CoV-2 variants

Bjarke Frost Nielsen a,b, Andreas Eilersen a, Lone Simonsen b, Kim Sneppen a,b,∗

a Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark
b Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark

A R T I C L E I N F O

Keywords:
Overdispersion
Evolution
Superspreading
Non-pharmaceutical interventions

A B S T R A C T

The SARS-CoV-2 ancestral strain has caused pronounced superspreading events, reflecting a disease characterized by overdispersion, where about 10% of infected people cause 80% of infections. New variants of the disease have different person-to-person variability in viral load, suggesting for example that the Alpha (B.1.1.7) variant is more infectious but relatively less prone to superspreading. Meanwhile, non-pharmaceutical mitigation of the pandemic has focused on limiting social contacts (lockdowns, regulations on gatherings) and decreasing transmission risk through mask wearing and social distancing. Using a mathematical model, we show that the competitive advantage of disease variants may heavily depend on the restrictions imposed. In particular, we find that lockdowns exert an evolutionary pressure which favours variants with lower levels of overdispersion. Our results suggest that overdispersion is an evolutionarily unstable trait, with a tendency for more homogeneously spreading variants to eventually dominate.

One of the major features of the coronavirus pandemic has been overdispersion in transmission, manifesting itself as superspreading. There is evidence that around 10% of infected individuals are responsible for 80% of new cases (Miller et al., 2020; Endo et al., 2020; Pozderac and Skinner, 2021; Kirkegaard and Sneppen, 2021). This means that some individuals have a high personal reproductive number, while the majority hardly infect at all. A recent study has shown this is reflected in the distribution of viral loads which is extremely wide, with just 2% of SARS-CoV-2 positive individuals carrying 90% of the virus particles circulating in communities (Yang et al., 2021). Overdispersion is in fact a key characteristic of certain diseases (Lloyd-Smith et al., 2005; Galvani and May, 2005; Woolhouse et al., 1997; Riley et al., 1962). However, this is by no means a universal signature of infectious respiratory diseases. Pandemic influenza, for example, is characterized by a much more homogeneous transmission pattern (Fraser et al., 2011; Brugger and Althaus, 2020; Roberts and Nishiura, 2011).

As an emerging virus evolves, its transmission patterns may change and it may become more or less prone to superspreading. The Alpha (B.1.1.7) variant of SARS-CoV-2 has been reported to be ∼50% more transmissible than the ancestral SARS-CoV-2 virus under varying degrees of lockdown (Graham et al., 2021; Volz et al., 2021; Davies et al., 2021). Meanwhile, others have shown that the Alpha variant possesses a higher average viral load and a reduced variability between infected persons, compared to the ancestral strain (Kidd et al., 2021). Preliminary analyses indicate that the Delta variant (B.1.617.2) is even less overdispersed (Mikszewski et al., 2022), while household transmission data suggest that the trend continues with the Omicron variant (B.1.1.529 and related lineages) (Jørgensen et al., 2022; Baker et al., 2022).

The altered viral load distributions seen in persons infected with the Alpha variant have also been investigated at the level of individual mutations. The spike protein of the Alpha variant prominently features the N501Y substitution (asparagine replaced by tyrosine at the 501 position) as well as the ΔH69/V70 deletion (histidine and valine deleted at the 69 and 70 positions). Investigators found that the viral load is,
Superspreading is a phenomenon that requires means (high biological infectiousness) as well as opportunity (social context). Such overdispersed transmission can have diverse origins, ranging from purely behavioural to biological (Althouse et al., 2020; Woolhouse et al., 1997). However, a recent meta-review (Chen et al., 2021) compared the transmission heterogeneity of influenza A (H1N1), SARS-CoV-1 and SARS-CoV-2 and found that higher variability in respiratory viral load was closely associated with increased transmission heterogeneity. This suggests that biological aspects of individual diseases are decisive in determining the level of overdispersion, and thus the risk of superspreading.

Across several diseases, individual variations in infectiousness have been approximated by a Gamma distribution (Lloyd-Smith et al., 2005) characterized by a certain mean value and a dispersion parameter known as $k$, which is related to the coefficient of variation ($CV$) through $CV = 1 / \sqrt{k}$. In the simplest of cases (a well-mixed population), infection attempts are modelled as a constant-rate (Poisson) process, which leads to a personal reproductive number which follows a negative binomial distribution. The dispersion parameter $k$ characterizes the degree of transmission heterogeneity; a lower $k$ corresponds to greater heterogeneity. For small values of $k$, it approximately corresponds to the fraction of infected individuals responsible for 80% of new infections. The value for the SARS-CoV-2 ancestral virus is around 10%, which is a priori obvious if a competitive advantage can be gained by specifically altering the variability in infectiousness (while keeping transmissibility unchanged). Our recent studies have shown that the presence of overdispersion makes a pandemic far more controllable than influenza pandemics when mitigating by limiting repetitive contacts (Sneppen et al., 2021) and personal contact network size (Nielsen et al., 2021). We therefore speculate that restrictions which alter social contact structure may, conversely, provide a fitness advantage to variants with more homogeneous transmission, and may thus play a role in viral evolution.

In this paper, we use a mathematical model to study the competition between idealized variants which differ in their level of overdispersion and their mean infectiousness. Our focus is on exploring whether overdispersion confers any evolutionary (dis)advantages, and whether non-pharmaceutical interventions which restrict social network size and transmissibility change the fitness landscape for variants with varying degrees of overdispersion. While it is evident that a higher mean infectiousness confers an evolutionary advantage to an emerging pathogen, it is not a priori obvious if a competitive advantage can be gained by specifically altering the variability in infectiousness. Our focus is on exploring whether overdispersion confers any evolutionary (dis)advantages, and whether non-pharmaceutical interventions which restrict social network size and transmissibility change the fitness landscape for variants with varying degrees of overdispersion. While it is evident that a higher mean infectiousness confers an evolutionary advantage to an emerging pathogen, it is not a priori obvious if a competitive advantage can be gained by specifically altering the variability in infectiousness (while keeping transmissibility unchanged). Our recent studies have shown that the presence of overdispersion makes a pandemic far more controllable than influenza pandemics when mitigating by limiting repetitive contacts (Sneppen et al., 2021) and personal contact network size (Nielsen et al., 2021). We therefore speculate that restrictions which alter social contact structure may, conversely, provide a fitness advantage to variants with more homogeneous transmission, and may thus play a role in viral evolution.

Across several diseases, individual variations in infectiousness have been approximated by a Gamma distribution (Lloyd-Smith et al., 2005) characterized by a certain mean value and a dispersion parameter known as $k$, which is related to the coefficient of variation ($CV$) through $CV = 1 / \sqrt{k}$. In the simplest of cases (a well-mixed population), infection attempts are modelled as a constant-rate (Poisson) process, which leads to a personal reproductive number which follows a negative binomial distribution. The dispersion parameter $k$ characterizes the degree of transmission heterogeneity; a lower $k$ corresponds to greater heterogeneity. For small values of $k$, it approximately corresponds to the fraction of infected individuals responsible for 80% of new infections. The value for the SARS-CoV-2 ancestral virus is around 10%, which is a priori obvious if a competitive advantage can be gained by specifically altering the variability in infectiousness (while keeping transmissibility unchanged). Our recent studies have shown that the presence of overdispersion makes a pandemic far more controllable than influenza pandemics when mitigating by limiting repetitive contacts (Sneppen et al., 2021) and personal contact network size (Nielsen et al., 2021). We therefore speculate that restrictions which alter social contact structure may, conversely, provide a fitness advantage to variants with more homogeneous transmission, and may thus play a role in viral evolution.

1. Initial survival of variants

The words fitness and competitive advantage may take on several meanings in an evolutionary context. For our purposes, it is especially important to distinguish between the ability of a pathogen to avoid stochastic extinction and to reproduce effectively in a population. To quantify the ability to avoid stochastic extinction we use a branching process to simulate an outbreak of a variant with a given level of overdispersion in a naive population. We then record whether it survives beyond the first 10 generations of infections, as a measure of the ability of that variant to take hold. Repeating these simulations multiple times allows us to compute the survival chance of each variant as a function of its infectiousness and overdispersion, in the absence and presence of mitigation (Fig. 1). Since we are dealing with a few related quantities, some definitions must be made. By the basic reproductive number ($R_e$) we mean the average number of new infections which each infected person gives rise to when all contacts are susceptible. This is in contrast to the effective reproductive number (known variously as $R_e$, $R_t$ and $R_0$), which is affected by population immunity, i.e. depletion of susceptibles. Note that $R_0$ as well as $R_e$ are context dependent, since behaviour (and mitigation strategies) will affect e.g. the number of contacts that a person has and thus the reproductive number. Another parameter entirely is the (biological) mean infectiousness, by which we mean the rate at which transmission occurs when an infected person is in contact with a susceptible person. This is a property of the disease and not of the social environment. In Fig. 1, the independent variables are thus the mean infectiousness and the dispersion parameter, both of which are assumed to be properties of the disease. The details of the calculation can be found in the Materials and Methods section.

In the unmitigated scenario (Fig. 1A), the procedure is relatively straightforward. A single infected individual is initially introduced,
with a personal reproductive number \( z \) drawn from a negative binomial distribution \( P(k|Z; R_0, k) \) with mean value \( R_0 \) and dispersion parameter \( k \). This individual then gives rise to \( z \) new cases, and the algorithm is recursively iterated for each subsequent infection.

In the case of a lockdown scenario, understood as a restriction on the number of social contacts (Fig. 1B), the algorithm is slightly more involved. In this case, a degree \( c \) (the number of contacts) is first drawn from a degree distribution (in this case a Poisson distribution, to mimic an Erdös–Renyi network). A biological reproductive number \( \xi \) (the infectiousness) is then drawn from a Gamma distribution with mean value \( R_0 \) and dispersion parameter \( k \). The actual personal reproductive number \( z \) is then drawn from the distribution

\[
P(z; \xi; c) = \frac{e^{-\xi/c} \cdot (\xi/c)^{z/c}}{z!}.
\]

This reflects that the personal reproductive number \( z \) is, naturally enough, limited by the number of distinct social contacts \( c \). This algorithm is then reiterated for each of the \( z \) new cases. In the limit of \( c \to \infty \), the distribution becomes Poissonian, recovering the homogeneous mixing limit as one would expect.

Similar results can be obtained analytically by viewing the spread as a Galton–Watson process and considering the probability that an infection chain dies out in infinite time. Denote that probability by \( d \) and let \( p_i \), \( i \in \{0, 1, \ldots\} \) be the distribution of personal reproductive numbers (i.e. \( p_i \) is the probability that a single infected individual will infect \( i \) others). Then the extinction risk \( d \) is given by the sum:

\[
d = p_0 + p_1 d + p_2 d^2 + \ldots
\]

where the first term on the right hand side is the extinction risk due to the index person producing no new infections, the second term corresponds to the case where the index person gives rise to one new branch which then dies out (this being the reason for the single factor of \( d \) in the second term) and so on. Since each new branch exists independently of the other, the extinction events are independent and the probabilities may be combined by simple multiplication as in (2).

Recognizing the quantity on the right hand side as the generating function \( G(d) \) for the distribution \( p_i \), the equation can be succinctly written as \( d = G(d) \). Note that this is, generically, a transcendental equation. The extinction probability \( d \), which appears on both sides of the equality, can thus not be algebraically solved for but must be determined through e.g. graphical or numerical means.

We find that the survival chance depends very strongly on overdispersion (Fig. 1), with more homogeneous variants (\( k \sim 1 \)) having a good chance of survival while highly overdispersed variants (\( k \leq 0.1 \)) are very unlikely to survive beyond 10 generations. This finding fits well with the general pattern of overdispersed spreading, namely that many individuals hardly become infectious while a few pass the disease onto many others. The uneven distribution of infectiousness makes heterogeneous diseases more fragile in the early stages of an epidemic, and thus more prone to stochastic extinction.

For the case of homogeneous mixing (Fig. 1A) and the number of generations tending to infinity, Lloyd-Smith et al. (Lloyd-Smith et al., 2005) performed a similar calculation using the generating function method described in (2). For a disease with \( R_0 = 3 \) and a \( k \) value of 0.16 (similar to what they estimated for SARS-CoV-1), the survival chance was found to be 24%. Our model yields the same figure in the unmitigated (connectivity→∞) limit.

To assess the effect of lockdown-like non-pharmaceutical interventions on the initial survival chances of a pathogen, we performed an analogous computation in a socially restricted setting (Fig. 1B). Compared with the unmitigated scenario of Fig. 1A, it can be seen that the mitigation has a differentiated effect on the survival chance, affecting highly overdispersed variants (small \( k \)) much more than their more homogeneous counterparts (despite the variants having the same mean infectiousness). This result is parallel to the effect of lockdown-like interventions on the competitive advantage of a variant, which we explore in the next section.

In Ref. Althouse et al. (2020), the authors study stochastic extinction of a superspreading disease under a targeted intervention they call cutting the tail. They introduce a cutoff value \( N_{\text{cutoff}} \) for the personal reproductive number, and if a person has a personal reproductive number \( z \geq N_{\text{cutoff}} \), a new \( z \) is drawn until one below the threshold is obtained. Since the disease is highly heterogeneous, this process is analogous to “removing” a potential superspreading event and replacing it with a much lower personal reproductive number (typically \( z = 0 \)). This is exactly why the intervention is rightly called targeted. Their approach is thus based on viewing superspreading entirely as an event-based phenomenon, where one can directly remove superspreading events above some threshold size, and instead let the individuals take part in other less risky events. Our approach, on the other hand, assumes superspreading to be due to a combination of high individual biological infectiousness and opportunity, e.g. a large number of social contacts.
These two viewpoints are complementary in obtaining a comprehensive description of superspreading phenomena, rather than mutually exclusive (Sneppen et al., 2021).

2. Competitive advantage is affected by context

We now turn to the competition between two variants which have already managed to gain a foothold, and so have moved past the initial risk of stochastic extinction. This is a separate aspect of “fitness”, distinct from the initial survival ability studied in the previous section. Fig. 2 explores the competition between two strains which differ only in their level of overdispersion. The ancestral variant has a broad infectiousness distribution ($k = 0.1$) while the other – the new variant – is more narrowly distributed ($k = 0.2$). Specifically, we probe how two such variants compete during a lockdown, as well as after the lockdown is lifted. In the initial partial lockdown scenario, each person is only allowed contact with 10 others. At first, the fraction of infections due to the new variant is observed to grow rapidly. When it reaches a 20% share of active infections, around day 65, the lockdown is lifted (simulated by a shift to a homogeneous mixing contact structure). Naturally, this more permissive contact structure causes a surge in both variants (Fig. 2c). However, the fraction of infections owing to each variant suddenly stabilizes, indicating that the more homogeneous new variant has lost its competitive advantage in the unmitigated scenario.

This sudden loss of competitive advantage demonstrates conceptually that the fitness of variants with different patterns of overdispersion depends on context, in the form of non-pharmaceutical interventions or the absence thereof. To quantify this dependence, we separately simulate the spread of several pathogen variants, each with its own specified mean infectiousness and dispersion parameter $k$, and measure the resulting basic reproductive numbers. In each case we use an individual based framework (see Materials and Methods) and let the pathogen spread in an Erdös–Renyi network with a mean connectivity of either 10 or 50, to simulate scenarios with either a restricted or fairly open society.

The results are shown in Fig. 3, where the competitive (dis)advantage of each variant is plotted as a function of its mean biological infectiousness and dispersion. The infectiousness is given relative to the SARS-CoV-2 ancestral strain which is set to average infectiousness $= 1$ and has dispersion $k = 0.1$. This average infectiousness of 1 corresponds to a basic reproduction number of $R_0 = 3$ in a well-mixed scenario, roughly representative of pre-Alpha SARS-CoV-2 (Billah et al., 2020). In the socially restricted case with only 10 contacts, the competitive advantage depends strongly on the dispersion parameter, as evidenced by the contour lines in Fig. 3A. The dashed white contour in the figure indicates variants which spread as well as the ancestral strain. Concretely, a variant with just half the biological infectiousness of the ancestral strain has no substantial competitive disadvantage, provided that it spreads sufficiently homogeneously ($k \geq 1.0$). In the more socially connected scenario (Fig. 3B), the competitiveness of a strain is observed to depend less strongly on dispersion, and is primarily determined by biological mean infectiousness. Viewed more broadly, these results imply that an observed increase in $R_0$ for an emerging variant may be due to a combination of changes in transmission patterns ($k$) and biological mean infectiousness.

The Erdös–Renyi contact network used in Fig. 3 has a relatively narrow (Poissonian) degree distribution, meaning that few high-degree nodes exist when the mean connectivity is low. However, even in a (partial) lockdown situation, some individuals – such as essential workers – may continue to have high connectivity. To probe the effects of this, we have performed the analysis of Fig. 3 in an approximately scale-free network based on findings by Danon et al. (2013). This type of network continues to have high-connectivity nodes, even at low mean connectivity. The result of this analysis, which is shown in Fig. S3 of the Supporting Information, is that a similar relation between overdispersion, biological transmissibility and fitness exists in such a scenario.

So far, our focus has been on mitigation strategies which rely on reductions in contact network size. However, even when societies reopen by allowing contact with an increased number of individuals, non-pharmaceutical interventions which decrease transmission risk per encounter may be in force. These may include face masks and regular testing. In the Supporting Information, we show that interventions which decrease the transmission risk per encounter (i.e. per unit of contact time) in fact decrease the competitive advantage of more homogeneous variants. These types of interventions thus have essentially the opposite effect, relative to strategies which reduce social connectivity.

3. Interventions exert selection pressure

As the observed differences in the viral load distributions of the Alpha (B.1.1.7) variant and the ancestral strain suggest, overdispersion is not a fixed property, but rather one that may evolve over time. Furthermore, SARS-CoV-2 has been estimated to mutate at a rate of approximately 2 substitutions per genome per month (Worobey et al., 2020), translating to about one mutation per three transmissions. In Fig. 4, we conceptually explore the consequences of overdispersion as an evolving feature of the pathogen. We utilize the same individual-based model as previously, but extend it with a simple model of trait mutation. In these simulations, the virus has a mutation probability of 1/3 at each transmission. When the pathogen mutates, the overdispersion factor is either increased (by a factor of 3/2) or decreased (by a factor of 2/3). Thus, we assume no drift on the microscopic scale.
but one may arise macroscopically due to selection pressure from the environment. It should be noted that while the assumed mutation rate is realistic for SARS-CoV-2, the frequency of mutations which affect the overdispersion trait is likely to be low, and as such the present model will likely overestimate the magnitude of the drift in overdispersion. It is however conceptually robust — decreasing the mutation rate merely slows down the drift, but the tendency remains.

In our simulations, we find that there is always a tendency for overdispersion to decrease (i.e. for the \( k \) value to increase), leading to more homogeneous disease transmission. This makes sense, since we have already established that heterogeneous disease variants are more likely to undergo stochastic extinction (Fig. 1) and that they have a competitive disadvantage as soon as contact structures are anything but well-mixed (Fig. 3). In the absence of any interventions, the tendency to evolve towards homogeneity is quite weak (Fig. 4A), but when a partial lockdown is instituted, the picture changes dramatically and the \( k \) value increases exponentially. The conclusion is thus that lockdowns exert a selection pressure on the virus when it comes to overdispersion, towards developing a less superspreading-prone phenotype.

One may of course object that the scenarios of Fig. 4A (unrestricted spread) and 4B (partial lockdown) are not directly comparable, since the epidemic in 4A unfolds much more rapidly. For this reason, we have included the scenario shown in 4C, where the transmission rate per encounter has been lowered, but social structure is unrestricted. The transmission rate is lowered such that the initial daily growth rates in Fig. 4B and 4C are identical (11%/day averaged over the first 14 days). This slightly increases the growth of \( k \) over the course of the epidemic, but to a much lower level than in the lockdown scenario, demonstrating that it is indeed the restriction of social network that provides the selection pressure driving \( k \) upwards.

4. Discussion

As demonstrated in this paper, the relative success and survival of mutants of a superspreading disease depends on the type of mitigation strategies employed within a population. The choice of a certain mitigation strategy may well amount to selecting the next dominant variant. If, for example, a simple lockdown is enacted while still allowing people to meet within restricted social groups, the evolution of more homogeneously spreading disease variants may become favoured.

That being said, our current findings should not be taken as advocating for any particular type of NPI over another — such recommendations require a much more comprehensive, multifactorial analysis. Rather, we wish to call for heightened attention to the role that behaviour in general — and NPIs in particular — can play in shaping pathogen evolution. Overdispersion is just one example of an evolving characteristic that may be affected by behaviour and population structure. Behavioural feedbacks on pathogen evolution comprise an emerging area of study, and other examples include the effects of population heterogeneity on the evolution of an asymptomatic first stage of infection (Saad-Roy et al., 2021) and of migration on influenza evolution (Bedford et al., 2010).

While our work shows that, all else being equal, certain NPIs may exert selective pressures via overdispersion, it is important to realize that other, stronger selective pressures may be present at any given time. For instance, a pressure towards higher biological transmissibility is generally expected. When a significant level of population immunity has been reached — through vaccination, recovery from infection or a combination thereof — strong selective pressures towards immune escape also become apparent. The relative strengths of these selective pressures are context dependent and are not easily ascertained.

The spreading of an emerging virus in a human society is a complex phenomenon, where the actual reproductive number depends on sociocultural factors, mitigation policies and self-imposed changes in the behaviour of citizens as awareness waxes and wanes in the population. The spread of a disease such as COVID-19 cannot simply be characterized by a single fitness quantity such as the basic reproductive number \( R_0 \), but will also depend on the heterogeneous transmission patterns within a population. If schools are open, mutants which spread more easily among children may be selected for, whereas rapid self-isolation of infected individuals may tend to favour variants which temporally separate disease transmission from the development of symptoms. We have focused on modelling the evolutionary effects of biological superspreading in the context of mitigations such as lockdowns which have been implemented globally during the COVID-19 pandemic. We find that such lockdowns tend to favour the emergence of more homogeneously spreading variants over time.

Our findings also have implications for the assessment of new variants. They highlight the importance of taking overdispersion into account when evaluating the transmissibility of an emerging variant. We have shown that a disease may spread more effectively not only by increasing its biological mean infectiousness, but also by changing its pattern of transmission to become more homogeneous. Practically, this means that transmission data obtained under even partial lockdown can lead to an over- or underestimation of the transmissibility of an emerging variant. We thus call for an increased focus on measurements the overdispersion of variants, as this may be critical for estimating the

Fig. 4. Evolution of overdispersion is driven by imposed restrictions. In these simulations, random mutations occur which alter the level of transmission overdispersion in a non-directed fashion. However, external evolutionary pressures are seen to drive the pathogen towards developing more homogenous spreading patterns. The filled red curve shows the combined incidence of all strains. The purple curve shows the average dispersion factor \( k \) in the infected population (with higher \( k \) corresponding to a more homogeneous infectiousness). The shaded purple area shows the 25% and 75% percentiles of the distribution of dispersion factors in the infected population. These simulations are run in an individual-based framework augmented with a simple evolutionary model, as detailed in the Materials and Methods. See Figure S4 in the Supporting Information for an analogous simulation in a scale-free network. See Figure S5 for the effect of varying the mutation rate and the magnitude of trait mutations. (A) The pathogen evolves in an open society with no restrictions imposed (homogeneous mixing contact structure). (B) Partial lockdown, with an average social network connectivity restricted to 15 persons. The social network is of the Erdös–Renyi type. (C) No restrictions on social network – a homogeneous mixing contact structure is assumed – but infectiousness is lowered by other means (e.g. face masks).
reproductive number of new variants SARS-CoV-2, and more generally in the monitoring of emerging pandemic threats.

5. Materials and methods

We use an individual-based (or agent-based) network model of disease transmission as originally developed in Ref. Nielsen et al. (2021). In this section, we present only a brief overview of the basic model, and refer to Ref. Nielsen et al. (2021) for a more detailed description. We then go on to describe in detail the simulations and calculations which are particular to this manuscript.

The disease progression model consists of four overall states, Susceptible, Exposed, Infected and Recovered. The exposed state has an average duration of 2.4 days and is subdivided into two consecutive states with exponentially distributed waiting times (i.e. having constant probability rate for leaving the state) of 1.2 days each, thus constituting a gamma distributed state when viewed as a whole. The infectious state is divided into two states as well, of 1.2 and 5 days in duration, respectively.

Each individual in the model is associated with a fixed social network. Only a subset of edges are activated in each timestep, to simulate a contact event. In the simulations of this work, we always use either an Erdős–Rényi network with finite mean connectivity, a scale free network (see Supporting Information) or a homogeneous-mixing contact structure. Homogeneous mixing is modelled by simply connecting pairs of randomly chosen individuals in each timestep. This is formally equivalent to the infinite connectivity limit of an Erdős–Rényi network as well.

When an edge connecting a susceptible and an infectious individual is active, there is a certain probability per unit of time for disease transmission to occur. This rate is determined by the individual infectiousness $r$, the infectious agent, which is drawn from a gamma distribution with dispersion parameter $k$ before the individual has become infectious. Such, the infectiousness for any given individual is assumed constant throughout the infectious stage of the disease. The infectiousness distribution determines an upper bound on size $\Delta t$ of the timesteps in the model, since the inequality $r \cdot \Delta t < 1$ must hold for all agents. A timestep of size $\Delta t = 30\text{min}$ was used throughout, since this was sufficient to ensure that the inequality was satisfied.

Below we go into more detail as to how the simulations involving multiple strains were performed.

5.1. Stochastic extinction

The stochastic extinction (or, conversely, survival) plots of Fig. 1 in the main text rely entirely on a branching process algorithm with sampling of probability distributions with an analytic description. In practice, we have performed the computational by numerical sampling.

In each generation of the epidemic, the computation is reiterated. For $i \in \{1, \ldots, I\}$:

- Draw individual infectiousness $\xi_i$ from Gamma distribution $P_i(\xi_i; k, \mu)$.
- Draw number of contacts $c$ from a Poisson distribution with a given mean connectivity.
- Given number of contacts $c$, draw personal reproductive number $z_i$ from the distribution (3)

$$P_i(z;\xi_i,c) = \binom{c}{z} (1 - e^{-\xi_i/c})^c \left(e^{-\xi_i/c} + e^{-(\xi_i + c)/c}\right)^{c-z}. \quad (3)$$

- Let the number of newly infected be $I = \sum_i z_i$ and repeat the algorithm with this new value of $I$.

If the number of infected $I$ ever drops to zero, the outbreak is said to have undergone stochastic extinction in that generation. By performing multiple such branching process simulations for each value of the parameters $\mu$ (mean infectiousness) and $k$ (dispersion factor) we build up a statistic of the survival chance of each specific variant. To generate Fig. 1, this is repeated for two different values of the mean connectivity $c$.

5.2. Two-strain competition simulations

In Fig. 2, two strains spread simultaneously in the population of $N = 10^8$ individuals. Initially, 0.099% of the population are infected with the heterogeneous "old" variant ($k = 0.1$), while 0.01% are infected with the more homogeneous "new" variant ($k = 0.2$). Once a person with a given variant infects a susceptible individual, the characteristics of the variant are passed on to the newly infected individual, such that the infectiousness of this person is drawn from a Gamma distribution with dispersion parameter $k$ set by the variant. In other words, these simulations assume that no further mutations affecting overdispersion occur, allowing us to track solely the competition of two differently-dispersed variants within a population.

5.3. Evolutionary model

In Fig. 4, we allow the pathogen to stochastically mutate upon transmission, with the mutations affecting the degree of overdispersion. In the simulations, the pathogen mutates once for every third transmission, on average (i.e. with mutation probability $p = 1/3$) and the mutations are assumed to always affect overdispersion, by either increasing the $k$ value by a factor of $3/2$ (i.e. $k \rightarrow 3k/2$) or decreasing it by a factor of $2/3$ (i.e. $k \rightarrow 2k/3$). On a microscopic level, the dispersion parameter thus performs an unbiased (multiplicative) random walk. The value of this step-size parameter is arbitrarily chosen, and as such the simulations can only be regarded as qualitative and conceptual. However, although no intrinsic bias is built into the mutation mechanism, external selection pressures may drive the level of overdispersion in the population up or down, as is explored in Fig. 4.

In Fig. 4C, the average infectiousness of the strain is lowered so as to produce an initial growth rate that is identical to that of 4A, namely 11% per day in the first 14 days of the epidemic.

CRediT authorship contribution statement

Bjarke Frost Nielsen: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing.
Andreas Eilersen: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing.
Lone Simonsen: Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing.
Kim Sneppen: Formal analysis, Funding acquisition, Investigation, Methodology, Software, Project administration, Resources, Supervision, Visualization, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Robert J. Taylor, and Julius B. Kirkegaard for enlightening discussions. Our research has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme, grant agreement No. 740704, as well as from the Carlsberg Foundation, Denmark under its Semper Ardens programme (grant # CF20-0046).
Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.epidem.2022.100613.

References


Kidd, Michael, Richter, Alex, Best, Angus, Cumley, Nicola, Mirza, Jeremy, Percival, Benita, Mayhew, Megan, Megram, Oliver, Ashford, Fiona, White, Thomas, et al., 2021. S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-qPCR. J. Infect. Dis. 223 (10).


