Phenotypic evolution of SARS-CoV-2: a statistical inference approach

Wakinyan Benhamou [iD], Sébastien Lion [iD], Rémi Choquet⋆ [iD] and Sylvain Gandon⋆ [iD]

CEFE, CNRS, Univ Montpellier, EPHE, IRD, Montpellier, France
⋆: equal contribution

August 25, 2022

Abstract

Since its emergence in late 2019, the SARS-CoV-2 virus has spread globally, causing the ongoing COVID-19 pandemic. In the fall of 2020, the Alpha variant (lineage B.1.1.7) was detected in England and rapidly spread and outcompeted the previous lineage. Yet, very little is known about the underlying modifications of the infection process that can explain this selective advantage. Here, we try to quantify how the Alpha variant differs from its predecessor on two phenotypic traits: the transmission rate and the duration of infectiousness. To this end, we analysed the joint epidemiological and evolutionary dynamics of SARS-CoV-2 as a function of the Stringency Index, a measure of the amount of Non-Pharmaceutical Interventions. We developed a two-step approach based on a *SEIR* model and the analysis of a combination of epidemiological and evolutionary information. First, we infer how the Stringency Index reduces the amount of viral transmission. Secondly, based on a novel theoretical derivation of the selection gradient in a *SEIR* model, we infer the phenotype of the Alpha variant from the analysis of the change in its frequency and we show that its selective advantage is more likely to result from a higher transmission than from a longer infectious period.

1 Introduction

In December 2019, acute pneumonias of as yet '*unknown etiology*' were increasingly reported in Wuhan, the capital of the Hubei Province in Central China [27]. Since then, the infectious agent responsible of this emerging zoonosis, a virus of the family *Coronaviridae* named SARS-CoV-2 (Severe Acute Respiratory Syndrome - CoronaVirus - 2), has spread worldwide, causing the pandemic COVID-19 (Coronavirus Disease-2019) [40] that is still ongoing today.

The possibility of a rapid SARS-CoV-2 adaptation was initially met with considerable scepticism [13, 34]. Indeed, compared to other single-stranded RNA viruses, the mutation rate of SARS-CoV-2 is

relatively low (estimated at the beginning of the pandemic around $6.8-9.8\times10^{-4}$ substitution.site^{-1}.year^{-1} [37, 38]). Besides, all the observed mutations in SARS-CoV-2 were initially thought to be neutral or slightly deleterious. The occasional rise of some mutations could be due to demographic stochasticity [12, 7] but the dramatic rise of specific mutations in different regions of the world challenged the hypothesis that none of these mutations were beneficial. In particular, the analysis of the emergence and the spread of several Variants of Concern (VOCs) across the world – e.g. Alpha (lineage B.1.1.7), Delta (lineage B.1.617.2) or Omicron (lineage B.1.1.529) (see for example `https://covariants.org/per-country` or `https://nextstrain.org/ncov/gisaid/global`) – demonstrated that these variants carry adaptive mutations that explain their faster rate of spread in the human population [28]. However, each of these mutations may act on various dimensions of the fitness landscape of the virus and affect different life-history traits. It is therefore much less clear *why* specific variants are favoured. In other words: which phenotypic trait(s) can explain this increase in viral fitness, i.e. the growth rate of the viral population ? Viral fitness is governed by multiple life-history traits like the transmission, the virulence or the recovery rates of the virus [7]. It is particularly important to understand which traits are involved in the increase in fitness because they may have very different implications for epidemiological dynamics and public health. For instance, an increase in the transmission rate or in the duration of infectiousness both lead to an increase in viral fitness but they may have distinct consequences for the efficacy of Non-Pharmaceutical Interventions (NPIs), implemented to mitigate the epidemic. It is therefore very important to understand and track this adaptation to optimize our control strategies.

In the following, we will focus on the first of these VOCs: the lineage B.1.1.7, categorised as *Variant of Concern 202012/01* and afterwards named "*Alpha variant*". This variant emerged in early fall 2020, in the South-East region of England [31, 39], and spread rapidly in the United Kingdom (**Fig. 1**) with an effective reproduction number estimated to be 40-100 % higher than for the previous lineage [3, 39]. In [3], the authors modelled several underlying biological mechanisms and suggested that a higher transmission rate per contact for the Alpha variant would be the most parsimonious explanation, but also that a longer duration of infectiousness – merely increasing the number of opportunities of transmission – would also explain the data very well. Besides, Otto *et al.* found that the selection coefficient of the Alpha variant (i.e. the slope of the change in its logit-frequency) was sensitive to the intensity of the control measures [29]. This intensity of control was taken into account using the "*Stringency Index*", a composite score that takes into account many response indicators such as lockdowns, school/workplace closures, face coverings or travel bans [14]. This index was computed since the beginning of the pandemic in many countries using data published by the Oxford COVID-19 Government Response Tracker (OxCGRT) [14]. As in [29], we found a negative correlation between the selection coefficient of the Alpha variant in England and the Stringency Index: -0.88 at the national level (95 % CI [-0.98; -0.39]) and between -0.97 (London, 95 % CI [-0.99; -0.86]) and -0.81 (South West, 95 % CI [-0.97; -0.14]) at the regional level (**Fig. S2**). Evolutionary epidemiology theory predicts that control measures that reduce contact rates with susceptible individuals will reduce the relative advantage of variants that have a transmission advantage [5, 7]. Hence, the negative effect of the Stringency Index on the selection coefficient suggests that the higher fitness of the Alpha variant is at least partly driven by higher transmission rates. Yet, it is difficult to rule out the possibility that

other phenotypic effects may be involved in this higher fitness [3].

The present work is an attempt to go a step further in the characterisation of the life-history traits of the Alpha variant. In particular we want to determine if the selective advantage is driven by a higher rate of transmission relative to the previous lineage, a higher duration of infectiousness (i.e. a lower recovery rate) or a combination of both. This new approach is based on a two-step analysis of two consecutive evo-epidemiological phases (**Fig. 1**). First, we infer the relationship between the Stringency Index and the effectiveness of the control measures (NPIs) from the analysis of the epidemiological dynamics taking place just before the emergence of the Alpha variant – i.e. just before the frequency of the variant reached 5 % of the cases tested positive in England (national level). Second, knowing the impact of NPIs on the viral propagation, we derived a novel theoretical expression for the selection coefficient of a variant in a $SEIR$ (Susceptible-Exposed-Infectious-Recovered) model and used it to infer the effects of the mutations of the Alpha variant on (i) the transmission rate and (ii) the duration of infectiousness from the analysis of the evolutionary dynamics taking place just after the emergence of the Alpha variant – i.e. just after the frequency of the variant exceeded 9.2 % of cases tested positive by region in England (regional level).

2 Results

In the first step of the analysis, our goal was to link the Stringency Index with the effectiveness of NPIs which reduces the contact with susceptible hosts. For this purpose, we developed a revised version of the $SEIR$ model (**Fig. S3**) formalised by a system of Ordinary Differential Equations (ODEs) (see equations (3) in **Methods, §4.1.1**). We used this model to generate numbers of daily new fatality cases (4), daily new cases tested negative (6) and daily new cases tested positive (7) that we compared to the observed data. We modeled the link between the effectiveness of the control measures $c(t)$ and the Stringency Index $\psi(t)$ for each time point t using the following function:

$$c(t) = k\left(\frac{\psi(t)}{100}\right)^{a} \tag{1}$$

with k, the maximum achievable effectiveness, and a, a "*shape*" parameter; $\psi(t)$ takes values between 0 (no control) and 100. For more details, see **Methods, §4.1.1**. The best Weighted Least Squares (WLS) estimator for this model yields $k = 1$ and $a = 3.78$. 95 % distributions, derived from Wild Bootstrap computations, of all the estimated parameters – including in particular parameters k and a – are shown in **Fig. S4**. Adjusted models seem to fit the general dynamics of the data even though somewhat locally perfectible (**Fig. S6**).

In the second step of the analysis we sought to explain the rapid spread of the Alpha variant through an increase in the transmission rate and/or the duration of infectiousness. We developed a $SEIR$ model which takes into account the circulation of both the Alpha variant and the previous lineage, which we will refer to as the resident strain (**Methods, §4.1.2**). This model was used to derive and approximation of the temporal dynamics of the frequency $\widetilde{f}_m(t)$ of the Alpha variant. Under the assumptions of weak selection and quasi-equilibrium of fast variables (for more details, see

SI Appendix), we obtained the following approximation for the selection coefficient of the Alpha variant:

$$s(t) = \frac{\mathrm{d}\ \mathrm{logit}(\widetilde{f}_m(t))}{\mathrm{d}t} \approx \frac{\kappa + \overline{r}(t)}{\kappa + \overline{\gamma}(t) + 2\overline{r}(t)} \left[\frac{\Delta\beta}{\overline{\beta}(t)} \left(\overline{r}(t) + \overline{\gamma}(t) \right) - \Delta\gamma \right], \tag{2}$$

with $\mathrm{logit}(\widetilde{f}_m(t)) = \ln\left(\widetilde{f}_m(t)/\big(1-\widetilde{f}_m(t)\big)\right)$ and where $\Delta\beta$ and $\Delta\gamma$ are the phenotypic differences between the Alpha variant and the resident strain in terms of transmission and recovery, respectively; $\overline{\beta}(t)$ and $\overline{\gamma}(t)$ refer to the average transmission and recovery rates across all genotypes; κ is the transition rate from the exposed state E to the infectious state I. Lastly, $\overline{r}(t)$ is the average growth rate of the epidemic:

$$\overline{r}(t) = q(t) \left((1 - c(t))\overline{\beta}(t) \frac{S(t)}{N} - \overline{\gamma}(t) \right),$$

with $q(t)$, the frequency of infectious individuals among infected hosts (i.e. $I(t)/\big(E(t)+I(t)\big)$). It is important to note that, in (2), NPIs affect the selection coefficient of the variant through the growth rate of the epidemic $\overline{r}(t)$ which depends on the amount of control $c(t)$. Crucially, this impact has a stronger effect on the term related to $\Delta\beta$ than the term related to $\Delta\gamma$. In the following, we approximated $\overline{r}(t)$ using the quasi-equilibrium expression of $q(t)$ and assuming that the ratio of susceptible hosts remained approximately constant during the second phase of the analysis ($S(t)/N \approx S/N$), and we neglected the effect of the rise in frequency of the variant on the average phenotypic trait values in (2) and $\overline{r}(t)$.

Under these assumptions, and using the previous best WLS estimates for parameters k and a from the first step of the analysis, a linear mixed-effects model (MEM) led to the following estimations of the phenotypic differences (per day): $\Delta\beta = 0.15$ (95 % CI $[0.033; 0.258]$) and $\Delta\gamma = -0.047$ (95 % CI $[-0.099; +0.001]$) (**Fig. 2**). With a significance level of 5 %, likelihood-based comparisons of nested MEMs show a significant effect for $\Delta\beta$ but not for $\Delta\gamma$ (although with a *p-value* very close to the significance threshold) (**Table S3**).

Additionally, using each of the 1999 pairs $\{k; a\}$ from Wild Bootstrap, we obtained as many estimates of $\Delta\beta$ and $\Delta\gamma$ such that 95 % of the estimates of $\Delta\beta$ were between 0.147 and 0.153 while 95 % of the estimates of $\Delta\gamma$ were between -0.054 and -0.046 (**Fig. S7**). Note that the CIs computed from each estimated value of $\Delta\beta \in [0.147, 0.153]$ remain positive. In contrast, the CIs computed from 61% of estimated value of $\Delta\gamma \in [-0.054, -0.046]$ cross the zero axis (**Fig. S8**). These distributions led us to conclude that the Alpha variant has a higher transmission rate than the resident strain. With these estimates of $\Delta\beta$ and $\Delta\gamma$, the selection coefficient $s(t)$ of the Alpha variant in the absence of NPI was computed, on average, around 0.77 per week (standard-deviation: 0.02 per week).

3 Discussion

We developed a two-step approach to characterise the phenotypic variation of the Alpha variant relative to the previously dominant lineage. In the first step of the analysis, we focus on the epidemiological dynamics before the emergence of the Alpha variant and we used an $SEIR$ model, a very simplified representation of an age-structured model, to infer the effect of the Stringency Index on the reduction

of transmission induced by these control measures. This led us to infer the convex increasing function that captures the effect of the Stringency Index on the reduction in the number of contacts with susceptible hosts (**Fig. S5**).

The second step of this approach is based on the analysis of the change in frequency of the Alpha variant after its emergence. Using evolutionary epidemiology theory [6, 5, 7], we derive an expression for the gradient of selection in a $SEIR$ model. The analysis of selection in such a class structured environment (i.e. the virus is infecting both the E and the I hosts) is facilitated with a weak selection approximation [24, 11, 25]. We recover a classical result derived in simpler SIR models: the intensity of selection for higher transmission rates depends on the availability of susceptible hosts and the amount of NPIs aiming to reduce contact (e.g. social distancing and face coverings). In contrast, selection for a longer duration of infectiousness is much less sensitive to these control measures. Using our independent estimation of the effectiveness of NPIs based on the Stringency Index, we inferred both $\Delta\beta$ and $\Delta\gamma$ of the Alpha variant from the temporal dynamics of its logit-frequency. This analysis shows that the selective advantage of the Alpha variant is mainly driven by a higher transmission rate and not by a longer duration of the infectiousness.

Several specific mutations of the Alpha variant could explain these phenotypic differences. Preliminary genomic characterisations detected around 17 non-synonymous substitutions or deletions compared to the previous lineage; about half were associated with the protein S gene, including mutations of immunological significance [32]. In particular, the mutation N501Y, known to increase the affinity of the viral glycoprotein S for the human receptor ACE2 [36], and the mutation P681H, adjacent to a serine protease cleavage site that is required for cell infection [19], are both likely to affect the within-host development of the virus in infected hosts. How this development affects key phenotypic traits like transmission and recovery rates in human host is difficult to explore experimentally. Our analysis can thus provide a complementary approach that may help to link genetic and phenotypic variation. Besides, after the completion of this study, viral shedding in breath aerosols were recently found to be higher in individuals infected with the Alpha variant than in those infected with previous strains [22].

Yet, it is important to note that this analysis relies on several simplifying assumptions. We assumed that infections start with an exposed but not yet infectious compartment ($SEIR$ model). We further assumed that, for the symptomatic cases, infectiousness began at the same time as the onset of symptoms – i.e. the latency period and the incubation period coincide perfectly in time. However, transmission from a pre-symptomatic state is a distinctive feature of SARS-CoV-2 [35]; it generally begins about 2.5 days before the onset of symptoms and is associated with the highest transmission in the life cycle of the virus (highest viral load) [7, 17]. Additionally, it would be particularly interesting to consider a more general model where the generation time could follow different distributions. This would allow us to consider alternative phenotypic dimensions like the investment in earlier transmission [4, 1, 16].

The availability of the data is a major limiting factor in any statistical inference analysis. More data and better quality data would of course allow us to go deeper in our analysis. The Stringency Index provides a rough approximation of the intensity of control. More precise and more local estimations of control would allow us to refine our estimations. In addition, we show in the **SI Appendix, §S6** how data frequency among different types of hosts (i.e. the differentiation between the exposed and

the infectious compartments) may provide another way to estimate $\Delta\beta$ and $\Delta\gamma$.

In spite of the above limits of our approach, we contend that it is important to attempt to exploit the joint epidemiological and evolutionary dynamics of SARS-CoV-2 to better understand its phenotypic evolution. This phenotypic evolution is undermining our efforts to control the epidemic. New variants are emerging and are affecting other phenotypic traits. In particular, the ability of new variants (e.g. Omicron) to escape immunity has major impact on the epidemiological dynamics [30]. Inference approaches using both epidemiological and evolutionary analysis could yield important insights on the adaptive trajectories on the phenotypic landscape of SARS-CoV-2.

4 Methods

4.1 A two-step analysis

The analysis is performed in two steps considering two consecutive evo-epidemiological periods of time: before and after the emergence of the Alpha variant in England (**Fig. 1**). The first step aims to estimate the force of infection in the presence of NPIs. In particular, we quantify $c(t)$, a function measuring the impact of NPIs at time t on the force of infection $\lambda(t)$. This first step takes place temporally before the emergence of the Alpha variant – i.e. before it reaches 5 % of the cases tested positive in England (national level) – and consists in modeling the epidemiological phase of the previous lineage, which we refer to as the resident strain, disregarding the pre-existing genetic diversity, e.g. D614G mutation [18]. The second step consists in estimating the differences in contagiousness and in infectious duration in the presence of NPIs during the period when the two strains cohabit – i.e. just after the frequency of the variant exceeds 9.2 % of cases tested positive by region in England (regional level). In addition to the Stringency Index, we also combine information from screening and mortality data for the first step (using an epidemiological model), while we focus on the changes in frequency of the variant among positive cases for the second step.

For both steps, we consider a host population of size N. We note S, E, I and R, respectively, the states (or compartments) of individuals that were Susceptible to the disease, Exposed – i.e. infected but not yet infectious -, Infectious and Recovered. For a given state, for instance S, we note $S(t)$ the density of people in that state at time t and $\dot{S}(t)$ its differentiation with respect to time. Let β be the *per capita* transmission rate (direct and horizontal) and γ the *per capita* recovery rate. Control measures implemented by governments – e.g. social distancing, lockdowns, travel bans – are NPIs that aim to curb the spread of the epidemic by alleviating the force of infection $\lambda(t) = \beta I(t)/N$. Given $c(t)$ the effectiveness of these measures – ranging from 0 (no control) to 1 (total control) – the expression for the force of infection thus becomes: $\lambda(t) = (1 - c(t))\beta I(t)/N$.

4.1.1 Step 1: epidemiological analysis just before the emergence of the Alpha variant

We use a version of the well-known $SEIR$ model (see **Fig. S3**) to estimate the parameters that govern the epidemiological dynamics before the arrival of the Alpha variant. We denote α, the additional *per capita* mortality rate induced by the viral disease (i.e. the virulence) and D, the compartment of deceased individuals (COVID-19-related). We assume that the (potential) onset of symptoms and

the beginning of infectiousness occur simultaneously after a latency period of mean duration $1/\kappa$. Within the infectious compartment I, some hosts develop symptoms (I_S) with probability ω while the others remain asymptomatic (I_A) with a complementary probability. It is further assumed that individuals I_A systematically recover at a *per capita* rate γ while individuals I_S are divided into two sub-compartments depending on their fate: I_{Sd}, with a probability p, for those who will eventually die from the disease, or, alternatively with a complementary probability, I_{Sr}, for those who will eventually recover (at the same rate as asymptomatic hosts).

To describe these epidemiological trajectories, we used the following system of ODEs:

$$\begin{cases} \dot{S}(t) &= -(1-c(t))\beta S(t)\dfrac{I(t)}{N} \\ \dot{E}(t) &= (1-c(t))\beta S(t)\dfrac{I(t)}{N} - \kappa E(t) \\ \dot{I}_A(t) &= (1-\omega)\kappa E(t) - \gamma I_A(t) \\ \dot{I}_{Sr}(t) &= (1-p)\omega\kappa E(t) - \gamma I_{Sr}(t) \\ \dot{I}_{Sd}(t) &= p\omega\kappa E(t) - \alpha I_{Sd}(t) \\ \dot{R}(t) &= \gamma\Big(I_A(t) + I_{Sr}(t)\Big) \\ \dot{D}(t) &= \alpha I_{Sd}(t) \end{cases} \tag{3}$$

Following [8] for the construction of the Next Generation Matrix, the basic reproduction number $\mathcal{R}_0$ – i.e. the expected number of *infectees* from one *infector* in a fully susceptible population – is then given in the absence of NPI by:

$$\mathcal{R}_0 = \beta\left(\frac{1-\omega p}{\gamma} + \frac{\omega p}{\alpha}\right).$$

In the context of COVID-19, the product ωp – i.e. the probability of developing symptoms and dying – is very low. We then approximated the basic reproduction number of the resident strain of SARS-CoV-2 as $\mathcal{R}_0 \approx \beta/\gamma$.

At each time point (each day), only a small fraction of the population is tested and hosts with symptoms are more likely to be tested than others. In order to take into account these biases, we used the following range of assumptions:

H1 - Individuals S and I_A are tested with the same probability / reporting rate ρ;
H2 - Individuals S and I_A can be tested several times;
H3 - All new individuals I_S (symptomatic) are tested (reporting rate of 1);
H4 - Screening of individuals E and R is neglected (reporting rate of 0);
H5 - All new disease-related deaths are reported (reporting rate of 1).

Furthermore, screening efforts in the UK have generally been strengthened over time during this period (as shown for instance by the increasing number of negative tests in **Fig. S6**). As the reporting rate for individuals without symptoms S and I_A can no longer be considered constant, we also assumed a linear increase with time:

H6 - The reporting rate ρ for individuals without symptoms S and I_A increases linearly over time: $\rho(t) = \eta\ t + \mu$.

The reporting rate is not identifiable in a SIR model when only a fraction of the compartment I are observed [15]. Thus, we also consider the disease-related deaths in the observation process. The combination of information, that is daily new cases tested negative, cases tested positive and fatality cases, allow us to identify the reporting rate.
Between two consecutive time points $t-1$ and t, the number of new fatality cases is given by:

$$D(t) - D(t-1) = \int_{t-1}^{t} \alpha I_{Sd}(t)\,\mathrm{d}t, \tag{4}$$

and, given $\int_{t-1}^{t} \omega\kappa E(t)\,\mathrm{d}t$, the *incidence* of symptomatic cases (i.e. new incomers in compartment I_S), we decomposed the number of performed tests $T(t)$ as follows:

$$T(t) = T^-(t) + T^+(t) = \underbrace{\rho(t)S(t)}_{T^-(t)} + \underbrace{\rho(t)I_A(t) + \int_{t-1}^{t} \omega\kappa E(t)\,\mathrm{d}t}_{T^+(t)}, \tag{5}$$

with $T^-(t)$ and $T^+(t)$, the number of cases tested negative and tested positive, respectively. Thus:

$$T^-(t) = (\eta\ t + \mu)\left[S(t-1) - \int_{t-1}^{t}\left((1-c(t))\beta S(t)\frac{I(t)}{N}\right)\,\mathrm{d}t\right] \tag{6}$$

$$T^+(t) = (\eta\ t + \mu)\left[I_A(t-1) + \int_{t-1}^{t}\left((1-\omega)\kappa E(t) - \gamma I_A(t)\right)\mathrm{d}t\right] + \int_{t-1}^{t} \omega\kappa E(t)\,\mathrm{d}t \tag{7}$$

Directly estimating the control efficiency $c(t)$ is usually impossible and $c(t)$ must then be modeled. However, it results from a multitude of factors that may vary spatially and temporally. Furthermore, $c(t)$ is not necessarily proportional to the severity of the measures in place. Here, we chose to use the *Stringency Index* (which we noted $\psi(t)$). With this score, the integration of many indicators is rescaled to a value between 0 (no control) and 100 (the stringest) in order to reflect the strictness of public health policy [14]. Although somewhat imperfect, this index has the advantage of integrating many factors into one value, as well as being available per day online since the beginning of the pandemic in many countries. To model the link between $c(t)$ and $\psi(t)$, we used the following concav or convex relationship:

$$c(t) = k\left(\frac{\psi(t)}{100}\right)^a \tag{1}$$

with $k \in [0;1]$, the maximum achievable efficiency – i.e. when $\psi(t) = 100$ –, and with $a \in \mathbb{R}_+^*$, a '*shape*' parameter.

4.1.2 Step 2: Evolutionary analysis

We considered now that two distinguishable infectious strains compete: the resident (or WT) strain, which will be represented with the subscript w, and the mutant strain (or variant), which will be

represented with the subscript m. The total number of exposed hosts $E(t)$ can therefore be decomposed into: $E(t) = E_m(t) + E_w(t)$. Likewise, for the infectious hosts $I(t)$: $I(t) = I_w(t) + I_m(t)$ and we denote $q_m(t) = I_m(t)/I(t)$, the frequency of the variant among I. We propose here that the variant may differ phenotypically from the resident strain in its effective transmission rate $\beta_m = \beta_w + \Delta\beta$ and/or its recovery rate $\gamma_m = \gamma_w + \Delta\gamma$. In contrast, we neglect any difference in terms of latency period ($\kappa_m = \kappa_w = \kappa$), and we neglect the virulence α of both strains. For SARS-CoV-2, frequencies of the Alpha variant do not seem to depend on the age of hosts [3]. Assuming furthermore that over-infections do not occur – including co-infections with both strains – and that (persistent) immunity acquired with either strain protects effectively against both, we used the simple following $SEIR$ model:

$$\begin{cases} \dot{S}(t) = -(1-c(t))\overline{\beta}(t)S(t)\dfrac{I(t)}{N} \\ \dot{E}(t) = (1-c(t))\overline{\beta}(t)S(t)\dfrac{I(t)}{N} - \overline{\kappa}(t)E(t) \\ \dot{I}(t) = \overline{\kappa}(t)E(t) - \overline{\gamma}I(t) \\ \dot{R}(t) = \overline{\gamma}(t)I(t) \end{cases} \tag{8}$$

where the overlines refer to mean values of the phenotypic traits after averaging over the distribution of strain frequencies:

$$\begin{cases} \overline{\beta}(t) &= (1-q_m(t))\beta_w + q_m(t)\beta_m \\ \overline{\gamma}(t) &= (1-q_m(t))\gamma_w + q_m(t)\gamma_m \end{cases}$$

As described in [24, 25], under the assumption of weak selection, the overall frequency of the variant at time t, $\widetilde{f}_m(t)$ can be tracked using:

$$\frac{\mathrm{d}\widetilde{f}_m(t)}{\mathrm{d}t} = \underbrace{\widetilde{f}_m(t)(1-\widetilde{f}_m(t))}_{\text{Genetic variance}} \underbrace{\mathbf{v}(t)^\top \Delta\mathbf{R}(t)\mathbf{f}(t)}_{\text{s(t), selection coefficient}} , \tag{9}$$

with $\mathbf{v}(t)$ the vector of reproductive values, $\mathbf{f}(t)$ the vector of class-frequencies within infected states, and $\Delta\mathbf{R}(t)$ the matrix of differences in transition rates between the mutant strain and the resident strain (for more details, see **SI Appendix**).

An easier way to study $s(t)$ in time series analyses is not to directly work with frequencies but with logit-frequencies instead, that is ln(frequency of the variant / frequency of the resident strain). Indeed, it may easily be shown that:

$$\frac{\mathrm{d}\ \mathrm{logit}(\widetilde{f}_m(t))}{\mathrm{d}t} = s(t) \tag{10}$$

We then focused on the selection coefficient of the variant $s(t)$. According to its value (weakly or strongly positive, weakly or strongly negative), this selection coefficient helps to explain the success or the disadvantage of the variant over the resident strain through natural selection [6, 5, 7]. In **SI Appendix**, we show that, using quasi-equilibrium approximations for fast variables, the selection

coefficient of the variant $s(t)$ may be approximated as:

$$s(t) \approx \frac{2\kappa(1-c(t))\Delta\beta\frac{S(t)}{N} - \left(\kappa - \overline{\gamma}(t) + \sqrt{\left(\kappa-\overline{\gamma}(t)\right)^2 + 4\kappa(1-c(t))\overline{\beta}(t)\frac{S(t)}{N}}\right)\Delta\gamma}{2\sqrt{\left(\kappa-\overline{\gamma}(t)\right)^2 + 4\kappa(1-c(t))\overline{\beta}(t)\frac{S(t)}{N}}} \tag{11}$$

For the SIR model nested in the above $SEIR$ model (8), the selection coefficient is merely: $s(t) = (1-c(t))\Delta\beta S(t)/N - \Delta\gamma$ [6, 5], which shows analytically the importance of $c(t)$ to distinguish the scenario where the variant has simply a higher transmission rate ($\Delta\beta > 0$; $\Delta\gamma = 0$) from the scenario where it only has a longer duration of infectiousness ($\Delta\beta = 0$; $\Delta\gamma < 0$), or from an intermediate scenario ($\Delta\beta \neq 0$; $\Delta\gamma \neq 0$). In other words, it is particularly the variations in this control that might help us to decouple the effects of these two phenotypic traits. Simply adding an exposed state E makes the selection gradient surprisingly much more difficult to express but the importance of the variations in $c(t)$ for this purpose (although less clear-cut) remains nevertheless relevant as suggested by (11).

4.2 Statistical inference

4.2.1 Programming

Digital simulations and data analyses were carried out using R (https://www.r-project.org/) version 4.1.1 (2021-08-10). ODEs were solved numerically by the function '*ode*' (method = '*ode45*') from the package '*deSolve*'.

4.2.2 Step 1

We used UK data between 2020-08-03 and 2020-11-08 (a period for which the Alpha variant was below 5 % among cases tested positive at national scale). Daily epidemiological data from national screening will be available in a future repository; 7-day rolling average data were used in order to mitigate the effects of variation in testing activity, e.g. during weekends. Daily COVID-19-related deaths in UK ('*Daily deaths with COVID-19 on the death certificate by date of death*') will be available in a future repository.

The goal of this part is to compute $c(t)$ from the Stringency Index and thus to focus on the estimation of the parameters k and a. We used additional information from the litterature to set the value of some parameters of the model. For instance, we set the mean latency period to 5 days [9] and the mean duration of infectiousness to 10 days [2] – that is, an average infection period of 15 days. Besides, we set the basic reproduction number $\mathcal{R}_0$ to 2.5 [10, 21, 23] and we approximated the initial states within compartment I. This is summarised with further details in **Table S1**. With these fixed parameter values, the model (3) is theoretically identifiable (**Fig. S10**, following [33]). Estimating remaining parameters for the first phase was performed using the method of Weigthed Least Squares (WLS). Let $\theta = \left(k, a, E(t_0^{\text{step 1}}), \alpha, \omega, p, \eta, \mu\right)^\top$ be the vector of parameters and $\widehat{\theta}$ its estimator. Observation states – here, (1) daily new cases tested negative, (2) daily new cases tested

positive and (3) daily new fatality cases – are specified with the subscript i and are modelled by functions f_i. Time points are noted t and y_i refers to the response variables. Hence:

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_i \sum_t \frac{\Big(y_{i,t} - f_i(\theta, t)\Big)^2}{f_i(\theta, t)}.$$

These non-linear optimizations were tackled with the R function '*optim*', from the basic package '*stats*', using the Nelder-Mead – or downhill simplex – method (maximum number of iterations $maxit = 2000$, absolute and relative convergence tolerance $abstol = reltol = 10^{-6}$). This optimization procedure was iterated for 1500 sets of initial values, because of the presence of local minima, and was restricted to certain ranges of values through parameter transformations (**Table S2**). Only parameter estimates from the best fit – i.e. with the lowest WLS value and successful completion – were kept. We refer to them in this text as the best WLS estimates.

Parameter distributions were then computed using Wild Bootstrap [26, 20], allowing in particular to take into account any heteroscedasticity in the residuals. To do this, 2000 sets of *bootstraped* data were generated and non-linear optimizations were reiterated, but starting only with the best WLS estimates and the corresponding set of initial values.

4.2.3 Step 2

Frequencies data were taken from the Public Health England (PHE) Technical Briefing 5, which was investigating the new VOC 202012/01 variant between September 2020 and January 2021 [31] (data will be available in a future repository). These data were reported weekly at the national scale as well as decomposed at the regional scale. Genotype frequencies resulted from qPCR performed after swab sampling in the the wider population – i.e. outside NHS hospitals and PHE labs. The qPCR tests from the ThermoFisher TaqPath kit target three genes: ORF1ab, N and S. However, the deletion ΔH69/V70 in the genome of the Alpha variant (in the gene encoding for the Spike glycoprotein S), results in a mismatch between the probe and the viral sequence. This lack of detection, referred to as SGTF (S-Gene Target Failure), could thus be used as a proxy for this very strain [31, 39]. In [31], a positive case was defined by the detection of the genes ORF1ab and N – i.e. with a cycle threshold of qPCR Ct that verifies $Ct \leq 30$ – while distinguishing between S-gene detection (Resident strain) or SGTF (Alpha variant) [31]. Eventually, the Stringency Index in the UK will be available in a future repository.

Under the assumption that variations in $S(t)/N$ on short time scales may be neglected for a controlled epidemic ($S(t)/N \approx S/N$) and by neglecting the effect of the rise in frequency of the variant on the average phenotypic trait values – i.e. $\overline{\beta}(t) = \beta_w$ and $\overline{\gamma}(t) = \gamma_w$ –, the explicit solution of equation (11) is then:

$$\begin{aligned}\operatorname{logit}(\widetilde{f}_m(t)) \approx \operatorname{logit}(\widetilde{f}_m(t_0^{\text{step 2}})) + \kappa \int_{t_0^{\text{step 2}}}^{t} \left(\frac{(1 - c(t))}{\sqrt{(\kappa - \gamma_w)^2 + 4\kappa(1 - c(t))\beta_w S/N}} \right) \mathrm{d}t \;\; \Delta\beta \frac{S}{N} - \\ \frac{1}{2} \left[(\kappa - \gamma_w) \int_{t_0^{\text{step 2}}}^{t} \left(\frac{1}{\sqrt{(\kappa - \gamma_w)^2 + 4\kappa(1 - c(t))\beta_w S/N}} \right) \mathrm{d}t + \Delta t \right] \;\; \Delta\gamma \end{aligned} \tag{12}$$

Where $\Delta t = t - t_0^{\text{step 2}}$ is the time between the system at time t and its initial state.

$\Delta\beta$ and S/N always appear as a product, which implies that they are not separately identifiable. Hence, in accordance with the end of our model fit for phase 1, we set S/N to 0.75. Besides, consistently with the first phase, we set: $\kappa = 0.2$, $\beta_w = 0.25$ and $\gamma_w = 0.1$. To estimate the phenotypic differences $\Delta\beta$ and $\Delta\gamma$ between the variant and the previous lineage in (12), we implemented a linear mixed-effects model (MEM) using the function '*lmer*' from the R package '*lme4*'. $\Delta\beta$ and $\Delta\gamma$ were treated as fixed effects and the region of England (9 in total) was treated as a random effect on the intercept of the model. Theses regions were considered independent of each other – i.e. we assumed no inter-region flows. For each region, the date of the initial state was the time when the Alpha variant variant outreached 9.2 % of cases tested positive – i.e. above horizontal lines in **Fig. S1-D**. Below this threshold, the dynamics of the variant could not be considered as deterministic. Note that the observed frequencies of the resident strain and the variant was considered as representative of the infected population. The parameters k and a that affect the link between Stringency Index and intensity of control were set according to their best WLS estimates and joint distribution that were previously computed in the first step (**§4.2.2**). 95 % CIs of parameters $\Delta\beta$ and $\Delta\gamma$ were computed using the function '*confint*' from the package '*stats*'.

Acknowledgements

We thank Troy Day and François Blanquart for inspiring discussions.

References

[1] Blanquart, F., Hozé, N., Cowling, B. J., Débarre, F., and Cauchemez, S. "Selection for infectivity profiles in slow and fast epidemics, and the rise of SARS-CoV-2 variants". *eLife* (2022). DOI: https://doi.org/10.7554/eLife.75791.

[2] Byrne, A. W., McEvoy, D., Collins, A. B., Hunt, K., Casey, M., Barber, A., Butler, F., Griffin, J., Lane, E. A., McAloon, C., O'Brien, K., Wall, P., Walsh, K. A., and More, S. J. "Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases". *BMJ Open* 10.8 (2020). DOI: https://doi.org/10.1136/bmjopen-2020-039856.

[3] Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., van Zandvoort, K., Silverman, J. D., CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz, K., Keogh, R., Eggo, R. M., Funk, S., Jit, M., Atkins, K. E., and Edmunds, W. J. "Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England". *Science* 372.6538 (2021). DOI: https://doi.org/10.1126/science.abg3055.

[4] Day, T. "Virulence evolution and the timing of disease life-history events". *TREE* 18.3 (2003). DOI: https://doi.org/10.1016/S0169-5347(02)00049-6.

[5] Day, T. and Gandon, S. "Applying population-genetic models in theoretical evolutionary epidemiology". *Ecology Letters* 10 (2007), 876–888. DOI: https://doi.org/10.1111/j.1461-0248.2007.01091.x.

[6] Day, T. and Gandon, S. "Insights From Price's Equation into Evolutionary Epidemiology". *Disease Evolution: Models, Concepts, and Data Analysis* (2006), 23–44.

[7] Day, T., Gandon, S., Lion, S., and Otto, S. P. "On the evolutionary epidemiology of SARS-CoV-2". *Curr. Biol.* 30.15 (2020), R841–R870. DOI: https://doi.org/10.1016/j.cub.2020.06.031.

[8] Diekmann, O., Heesterbeek, J. A. P., and Roberts, M. G. "The construction of next-generation matrices for compartmental epidemic models". *J.R. Soc. Interface* 7.47 (2010), 873–885. DOI: https://doi.org/10.1098/rsif.2009.0386.

[9] Ding, Z., Wang, K., Shen, M., Wang, K., Zhao, S., Song, W., Li, R., Li, Z., Wang, L., Feng, G., Hu, Z., Wei, H., Xiao, Y., Bao, C., Hu, J., Zhu, L., Li, Y., Chen, X., Yin, Y., Wang, W., Cai, Y., Peng, Z., and Shen, H. "Estimating the time interval between transmission generations and the presymptomatic period by contact tracing surveillance data from 31 provinces in the mainland of China". *Fundamental Research* 1.2 (2021), 104–110. DOI: https://doi.org/10.1016/j.fmre.2021.02.002.

[10] Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L. C., van Elsland, S., Thompson, H., Verity, R., Volz, E., Wang, H., Wang, Y., Walker, P. G., Walters, C., Winskill, P., Whittaker, C., Donnelly, C. A., Riley, S., Ghani, A. C., and on behalf of the Imperial College COVID-19 Response Team. "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand" (2020). URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf.

[11] Gandon, S. and Lion, S. "Targeted vaccination and the speed of SARS-CoV-2 adaptation". *PNAS* 119.3 (2022). DOI: https://doi.org/10.1073/pnas.2110666119.

[12] Grubaugh, N. D., Hanage, W. P., and Rasmussen, A. L. "Making Sense of Mutation: What D614G Means for the COVID-19 Pandemic Remains Unclear". *Cell* 182 (2020), 794–795. DOI: https://doi.org/10.1016/j.cell.2020.06.040.

[13] Grubaugh, N. D., Petrone, M. E., and Holmes, E. C. "We shouldn't worry when a virus mutates during disease outbreaks". *Nat. Microb.* 5 (2020), 529–530. DOI: https://doi.org/10.1038/s41564-020-0690-4.

[14] Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., and Tatlow, H. "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)". *Nat. Hum. Behav* 5 (2021), 529–538. DOI: https://doi.org/10.1038/s41562-021-01079-8.

[15] Hamelin, F., Iggidr, A., Rapaport, A., and Sallet, G. "Observability, Identifiability and Epidemiology A survey". *HAL* (2021). URL: https://hal.archives-ouvertes.fr/hal-02995562/document.

[16] Hart, W. S., Abbott, S., Endo, A., Hellewell, J., Miller, E., Andrews, N., Maini, P. K., Funk, S., and Thompson, R. N. "Inference of the SARS-CoV-2 generation time using UK household data". *eLife* (2022). DOI: https://doi.org/10.7554/eLife.70767.

[17] He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B. J., Li, F., and Leung, G. M. "Temporal dynamics in viral shedding and transmissibility of COVID-19". *Nat. Med.* 26 (2020), 672–675. DOI: https://doi.org/10.1038/s41591-020-0869-5.

[18] Hodcroft, E. B., Zuber, M., Nadeau, S., Vaughan, T. G., Crawford, K. H. D., Althaus, C. L., Reichmuth, M. L., Bowen, J. E., Walls, A. C., Corti, D., Bloom, J. D., Veesler, D., Mateo, D., Hernando, A., Comas, I., González-Candelas, F., SeqCOVID-SPAIN consortium, Stadler, T., and Neher, R. A. "Spread of a SARS-CoV-2 variant through Europe in the summer of 2020". *Nature* 595 (2021), 707–712. DOI: https://doi.org/10.1038/s41586-021-03677-y.

[19] Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. "A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells". *Mol. Cell.* 78.4 (2020), 779–784. DOI: https://doi.org/10.1016/j.molcel.2020.04.022.

[20] Kline, P. M. and Santos, A. "A Score Based Approach to Wild Bootstrap Inference". *J. Econom. Methods* 1.1 (2012), 23–41. DOI: https://doi.org/10.1515/2156-6674.1006.

[21] 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 (2020), 553–558. DOI: https://doi.org/10.1016/S1473-3099(20)30144-4.

[22] Lai, J., Coleman, K. K., Sheldon Tai, S.-H., German, J., Hong, F., Albert, B., Esparza, Y., Srikakulapu, A. K., Schanz, M., Sierra Maldonado, I., Oertel, M., Fadul, N., Louie Gold, T., Weston, S., Mullins, K., McPhaul, K. M., Frieman, M., and Milton, D. K. "Evolution of SARS-CoV-2 Shedding in Exhaled Breath Aerosols". *MedRxiv* (2022). DOI: https://doi.org/10.1101/2022.07.27.22278121.

[23] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T. Y., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., and Feng, Z. "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia". *N. Engl. J. Med.* 382.13 (2020), 1199–1207. DOI: https://doi.org/10.1056/NEJMoa2001316.

[24] Lion, S. "Class structure, demography and selection: reproductive-value weighting in nonequilibrium, polymorphic populations". *Am. Nat.* 191.5 (2018). DOI: https://doi.org/10.1086/696976.

[25] Lion, S. and Gandon, S. "Evolution of class-structured populations in periodic environments". *bioRxiv* (2022). DOI: https://doi.org/10.1101/2021.03.12.435065.

[26] Liu, R. Y. "Bootstrap Procedures under some Non-I.I.D. Models". *Ann. Stat.* 16.4 (1988), 1696–1708. DOI: https://doi.org/10.1214/aos/1176351062.

[27] Lu, H., Stratton, C. W., and Tang, Y.-W. "Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle". *J. Med. Virol.* 92.4 (2020), 401–402. DOI: https://doi.org/10.1002/jmv.25678.

[28] Obermeyer, F., Jankowiak, M., Barkas, N., Schaffner, S. F., Pyle, J. D., Yurkovetskiy, L., Bosso, M., Park, D. J., and Babadi, M. "Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness". *Science* (2022). DOI: https://doi.org/10.1126/science.abm1208.

[29] Otto, S. P., Day, T., Arino, J., Colijn, C., Dushoff, J., Li, M., Mechai, S., Van Domselaar, G., Wu, J., Earn David, J. D., and Ogden, N. H. "The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic". *Curr. Biol.* 31.14 (2021). DOI: https://doi.org/10.1016/j.cub.2021.06.049.

[30] Paton, R. S., Overton, C. E., and Ward, T. "The rapid replacement of the Delta variant by Omicron (B. 1.1. 529) in England". *Sci. Transl. Med.* (2022), eabo5395. DOI: https://doi.org/10.1126/scitranslmed.abo5395.

[31] Public Health England. "Investigation of novel SARS-COV-2 variant 202012/01: technical briefing 5" (2020). URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959426/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5.pdf.

[32] Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D. L., Volz, E., and on behalf of COVID-19 Genomics Consortium UK (CoG-UK). "Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations". *Virological* (2020). URL: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563.

[33] Raue, A., Kreutz, C., Maiwald, T., Bachman, J., Schilling, M., Klingmüller, U., and Timmer, J. "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood". *Bioinformatics* 25.15 (2009), 1923–1929. DOI: https://doi.org/10.1093/bioinformatics/btp358.

[34] Rausch, J. W., Capoferri, A. A., Katusiime, M. G., Patro, S. C., and Kearney, M. F. "Low genetic diversity may be an Achilles heel of SARS-CoV-2". *PNAS* 117.40 (2020), 24614–24616. DOI: https://doi.org/10.1073/pnas.2017726117.

[35] Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., Seilmaier, M., Drosten, C., Vollmar, P., Zwirglmaier, K., Zange, S., Wölfel, R., and Hoelscher, M. "Transmission of 2019-nCoV Infection from an Asymptomatic

Contact in Germany". *N. Engl. J. Med.* 382 (2020), 970–971. DOI: https://doi.org/10.1056/NEJMc2001468.

[36] Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veesler, D., and Bloom, J. D. "Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding". *Cell* 182.5 (2020), 1295–1310. DOI: https://doi.org/10.1016/j.cell.2020.08.012.

[37] van Dorp, L., Richard, D., Tan, C. C. S., Shaw, L. P., Acman, M., and Balloux, F. "No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2". *Nat. Commun.* 11.5986 (2020). DOI: https://doi.org/10.1038/s41467-020-19818-2.

[38] Vasilarou, M., Alachiotis, N., Garefalaki, J., Beloukas, A., and Pavlidis, P. "Population Genomics Insights into the First Wave of COVID-19". *Life* 11.129 (2021). DOI: https://doi.org/10.3390/life11020129.

[39] Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D. K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D. P., The COVID-19 Genomics UK (COG-UK) consortium, Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Andrew, R., and Ferguson, N. M. "Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England". *Nature* 593 (2021), 266–269. DOI: https://doi.org/10.1038/s41586-021-03470-x.

[40] World Health Organization (WHO). "WHO Situation Report on 11 February 2020". 22 (2020). URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2.

Abbreviations (alphabetical order)

ACE2	:	Angiotensin-Converting Enzyme 2
CI	:	Confidence Interval
COVID-19	:	COronaVIrus Disease 2019
i.i.d.	:	independent and identically distributed
MEM	:	Mixed-Effects Model
NHS	:	National Health Service (UK)
NPI	:	Non-Pharmaceutical Intervention
ODE	:	Ordinary Differential Equation
ORF1ab	:	Open Reading Frames 1a and 1b
OxCGRT	:	Oxford COVID-19 Government Response Tracker
PHE	:	Public Health England
qPCR	:	quantitative Polymerase Chain Reaction
S	:	Spike
SARS-CoV-2	:	Severe Acute Respiratory Syndrome - CoronaVirus - 2
$\boldsymbol{S(E)IR}$	:	Susceptible-(Exposed-)Infectious-Recovered
SGTF	:	S-Gene Target Failure
VOC	:	Variant Of Concern
WHO	:	World Health Organization
WLS	:	Weighted Least Squares
WT	:	Wild Type

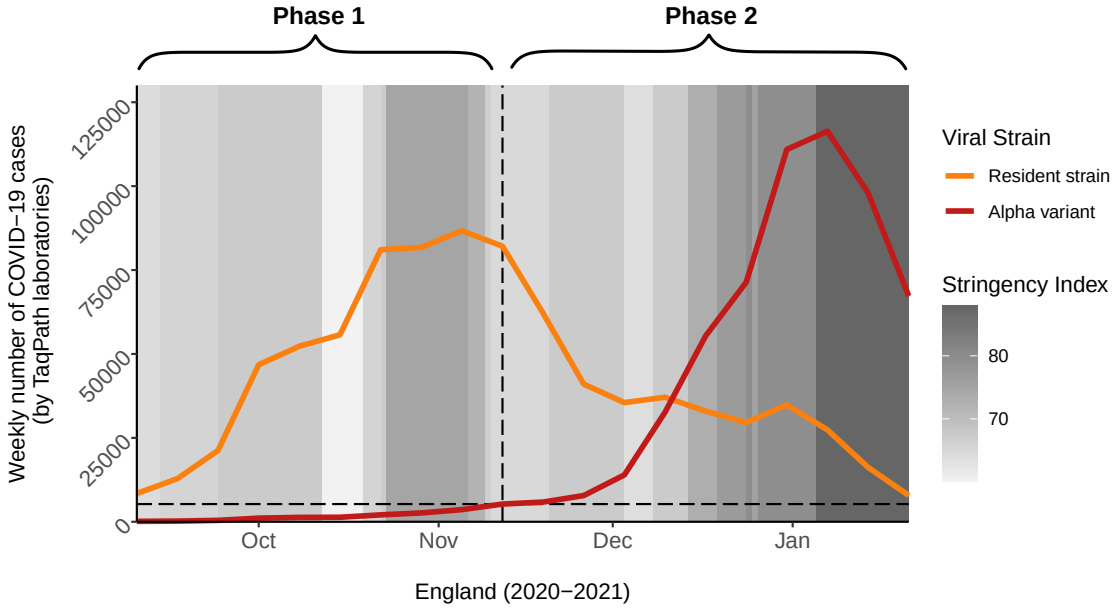

Figure 1: **The two consecutive phases of the spread of the Alpha variant.** In phase 1 – before the emergence of the Alpha variant –, we assume the epidemic is driven solely by the resident strain; in phase 2 – after the emergence of the Alpha variant, – we assume the epidemic results from the joint dynamics of the resident strain and the Alpha variant. In the first step of our analysis we estimated the impact of the Stringency Index – a measure of the amount of NPIs implemented to mitigate the epidemic, from 0 (no control) to 100 (stringest control) – on the propagation of the resident strain during phase 1. In the second step of our analysis, knowing the impact of NPIs, we estimated the phenotypic differences between the resident strain and the Alpha variant during phase 2. The dates reported on the chart match the middle of each week (Thursday). We set the end of phase 1 at the date when the Alpha variant reached 5 % of the cases tested positive at the national level (horizontal dashed line). For the sake of simplicity, we show the data at the national level but the starting date of phase 2 varied among regions (see **Methods, §4.1** and **Fig. S1**).

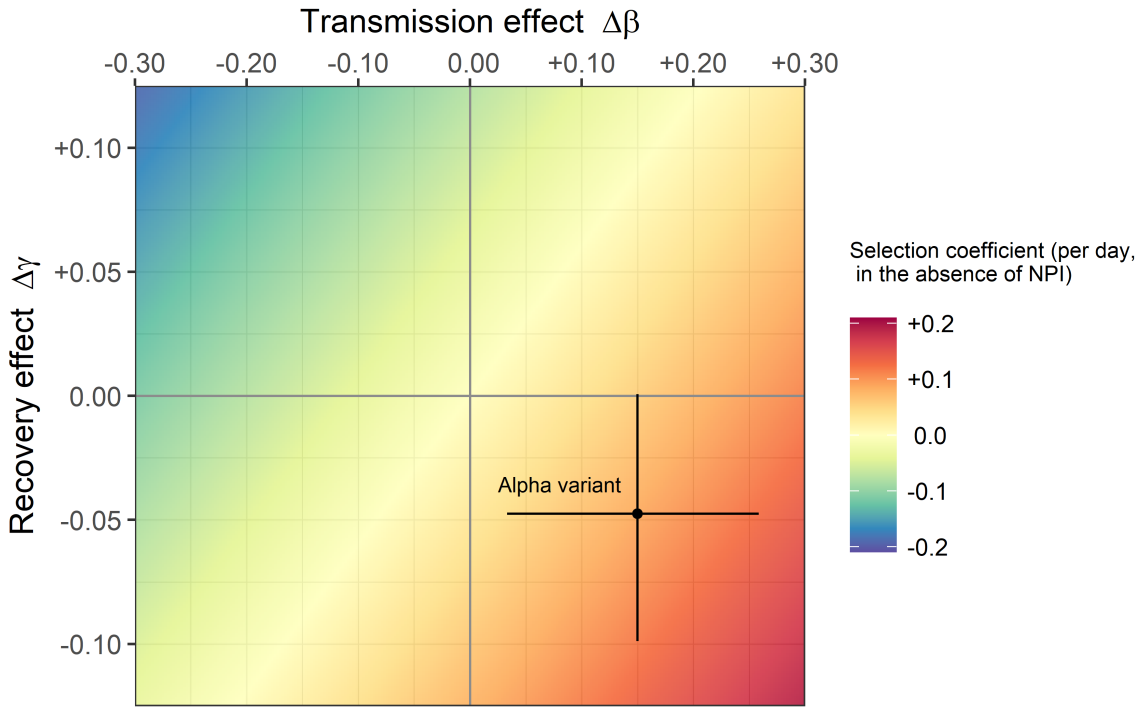

Figure 2: **Alpha variant phenotypes (transmission and recovery rates) relative to the resident strain.** Linear MEM estimates of phenotypic differences per day in transmission $\Delta\beta$ and in recovery $\Delta\gamma$ (black point) and 95 % CIs (black cross) are based on the best WLS estimates of control parameters k and a from the analysis of phase 1 ($k = 1$ and $a = 3.78$). We obtained $\Delta\beta = 0.15$ (95 % CI $[0.033; 0.258]$) and $\Delta\gamma = -0.047$ (95 % CI $[-0.099; +0.001]$). Other parameter values are $S/N = 0.75$, $\kappa = 0.2$, $\beta_w = 0.25$ and $\gamma_w = 0.1$. Estimates and 95 % CIs based on the joint distributions of parameters k and a from Wild Bootstrap computations are represented in **Fig. S8**. By definition, the phenotype of the resident strain is located at the origin of the graph ($\Delta\beta = 0$; $\Delta\gamma = 0$). The colored background represents the values of the selection coefficient as a function of $\Delta\beta$ and $\Delta\gamma$ (in the absence of NPI). Within this framework, the estimates of the phenotypic differences (per day) $\Delta\beta = 0.15$ and $\Delta\gamma = -0.047$ yield a selection coefficient $s(t)$ of the Alpha variant of 0.11 per day (thus, 0.77 per week).