ELSEVIER

Short communication

A very simple model to account for the rapid rise of the alpha variant of SARS-CoV-2 in several countries and the world

Hugo Fort

Institute of Physics, Faculty of Science, Universidad de la República, Iguá 4225, Montevideo 11400 Uruguay

A B S T R A C T

Since its first detection in the UK in September 2020, a highly contagious version of the coronavirus, the alpha or British variant a.k.a. B.1.1.7 SARS-CoV-2 virus lineage, rapidly spread across several countries and became the dominant strain in the outbreak.

Here it is shown that a very simple evolutionary model can fit the observed change in frequency of B.1.1.7 for several countries, regions of countries and the whole world with a single parameter, its relative fitness f, which is almost universal $f \approx 1.5$. This is consistent with a 50% higher transmissibility than the local wild type and with the fact that the period in which this variant takes over has been in all the studied cases around 22 weeks.

In September 2020 a new variant of the SARS-CoV-2 virus, known as lineage B.1.1.7 (aka 20I/501Y.V1, Variant of Concern 202012/01, British or Alpha variant), was detected in the UK and it quickly displaced the other variants of this virus in several countries (Hodcroft 2021).

Previous estimates of B.1.1.7 transmissibility have varied across different studies. For example: a preprint by Davis et al. (2020) reported that the variant was 56% (50%–74%) more transmissible than other variants across three regions in England (East of England, South East of England, and London), while another article concluded that it was 75% (70%–80%) more transmissible in the UK between October and November 2020 (Leung et al. 2021).

Here a very simple evolutionary model is proposed for describing the change in frequency of B.1.1.7. This model distinguishes the B.1.1.7 variant from all the others (treated just as one 'mean' lineage). If we call x the fraction of the B.1.1.7 variant a simple equation to model evolution by natural selection is given by (Nowak 2006, chapter 2):

$$\frac{dx}{dt} = x\left(f - \bar{f}\right), \tag{1}$$

where f is the fitness of this lineage relative to the others (which are set to 1) and $\bar{f}$ is the weighted mean fitness over all the variants, i.e.

$$\bar{f} = x \times f + (1-x) \times 1 = x(f-1) + 1. \tag{2}$$

Substituting (2) into (1) we obtain:

$$\frac{dx}{dt} = (f-1)x(1-x), \tag{3}$$

i.e. a logistic equation with a growth rate of $f-1$. Eqs. (1) to (3) are well

known equations to model natural selection. The solution of (3) is also well known and is given by (Fort 2020):

$$x(t) = \frac{1}{1 + \left(\frac{1}{x(0)} - 1\right)e^{-(f-1)t}}. \tag{4}$$

For long times, if $f > 1$, x will converge to an asymptotic value $x = 1$.

Then, Eq. (4) is compared against empirical data for those countries such that a minimum statistics of 500 sequences in each measurement were included (Table 1). That is, countries are considered if they have at least $N_{min} = 500$ sequences summing, at each time, all the variants that were tracked since the first time the B.1.1.7 lineage was reported in each country (Hodcroft 2021).

It turns out that this formula can fit the observed change in frequency x of B.1.1.7 until it became 1 or close to 1 for several countries, regions of countries with a single parameter, the relative fitness f of this variant, which is almost universal. It is worth remarking that in this communication we are only addressing the "rise" of B.1.1.7. Its "fall", which already occurred in countires like UK, involves the effect of factors like massive vaccination and new non-native variants of coronavirus (e.g. the Delta variant) which lie beyond this simple modeling. This modeling is also compared for the whole world, where x has decreased from a peak of $x \approx 0.77$, reached the 5th of May 2021, to below 0.2 since the 8th of June 2021 (Latif 2021).

Thus in each case I varied the parameter f in steps of 0.1 and chosed the one that produced minimum mean absolute error (MAE).

Fig. 1 shows how this fraction x has grown in the four countries of Table 1 and the comparison with the values predicted by Eq. (4) with the time t measured in weeks. Notice that this formula in general fits very

Table 1
Data for the countries included in this study. N_{min} stands for the minimum total number of sequences (not cases) measured (see text).

Country	N_{min} across weeks	Period	Relative fitness f
Denmark[1]	519	2020-02-26 to 2021-05-03	1.54
Netherlands[2]	500	2020-07-12 to 2021-05-03	1.45
Switzerland[2]	883	2020-07-12 to 2021-05-03	1.53
UK[2]	8217	2020-16-10 to 2021-03-22	1.54

[1] Danish Covid-19 2021
[2] Hodcroft 2021

Table 2
Goodness of the fit.

	Denmark	Netherlands	Switzerland	UK	World
degrees of freedom n-1	23	11	10	11	12
χ^2	2.271	4.061	4.650	1.656	7.757
pV	0.999	0.968	0.913	0.998	0.817
MAE	0.023	0.033	0.059	0.030	0.034

well with the data, yielding a clean sigmoid curve when plotted vs. time. This would be expected under controlled laboratory conditions, but not so much for this kind of measurements involving several non controlled factors and different kind of heterogeneities.

In the case of Switzerland (panel d) the frequency of B.1.1.7 reached a maximum of $x = 0.9$ and then started to decline. So I also included a fit with an asymptotic carrying capacity K = 0.9 (dashed line) just to show that the logistic fit for the last part of the sequence greately improves.

Interestingly, the best fit for the four countries occurs for similar values of the relative fitness parameter; around 1.5. Additionally, the values of $f =1.54$ and $f =1.53$ for, respectively, Denmark and Switzerland are in agreement with a reported 52% higher transmissibility when compared to the wildtype in Denmark and a 51% higher transmissibility when compared to the wildtype in Switzerland (ISPM 2021). This suggests that it is right to interpret f as the relative fitness to the local wildtype.

It seems interesting to find out if the model still works for smaller spatial scales, for instance across different regions of a country. In fact, in many countries, sampling may not be equal across the country: samples may only cover one area or certain areas (Hodcroft 2021). Fig. 2 shows the data for the five different regions of Denmark. Notice that, except for Nordjylland, the model fits the data quite well.

Finally, let us see now the application of the model to the maximum spatial scale, i.e. the whole world. Fig. 3 shows the empirical data for the world (Latif 2021) and the fit provided by Eq. (4) at the beginning is not very good but then it yields again a remarkable agreement with the empirical data for a smaller value of the relative fitness. This failure for the initial weeks can be understood since the pace of the pandemic varied for countries across the world as well as the quantity and quality of measurements (mainly for the initial stages in the rise of B.1.1.7).

I computed two metrics to quantitatively assess the quality of the fit: the pValue of a χ^2 goodness of fit test and the MAE. Table 1 shows these results. Notice tha both indices show the fitting is relevant. In particular the MAE varies from 0.023 to 0.059, which for a variable like x in [0,1], corresponds to a quite small average relative error (between 4.6 and 11.8 %).

In summary:

1 The change in frequency of SARS-CoV-2 virus lineage B.1.1.7 in time is well described by formula (4) resulting from an elementary evolutionary model with a single parameter, the relative fitness f.
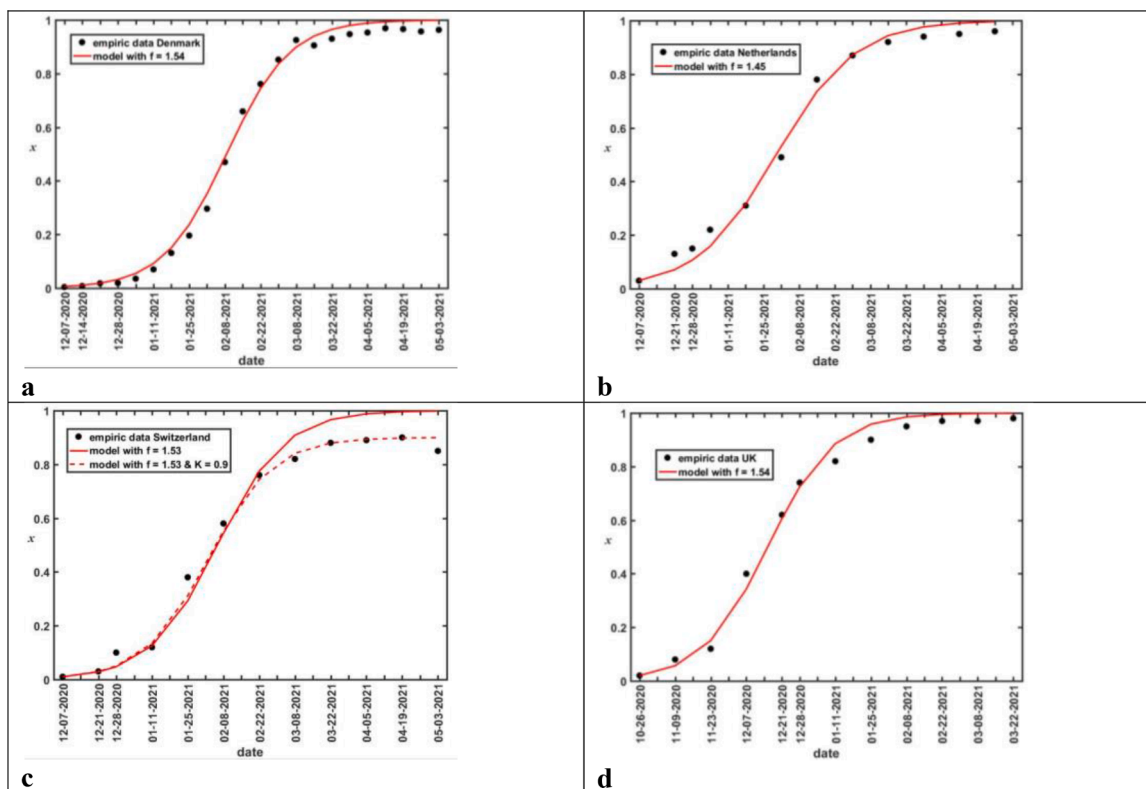
Figure 1. Change in the frequency of B.1.1.7 in the four countries of Table 1; empirical data (black circles) and formula (4) (solid line). (a) Denmark. (b) Netherlands. (c) Switzerland. (d) UK.
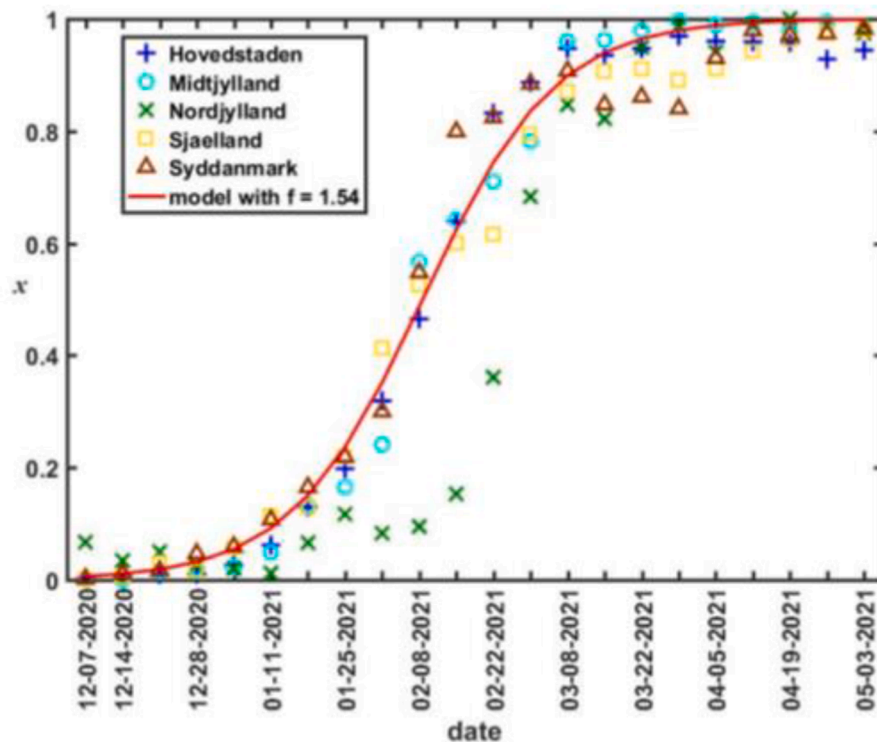
Figure 2. Change of the frequency of B.1.1.7 for different regions of Denmark as a percentage, empirical data (symbols) and formula (4) (red line).
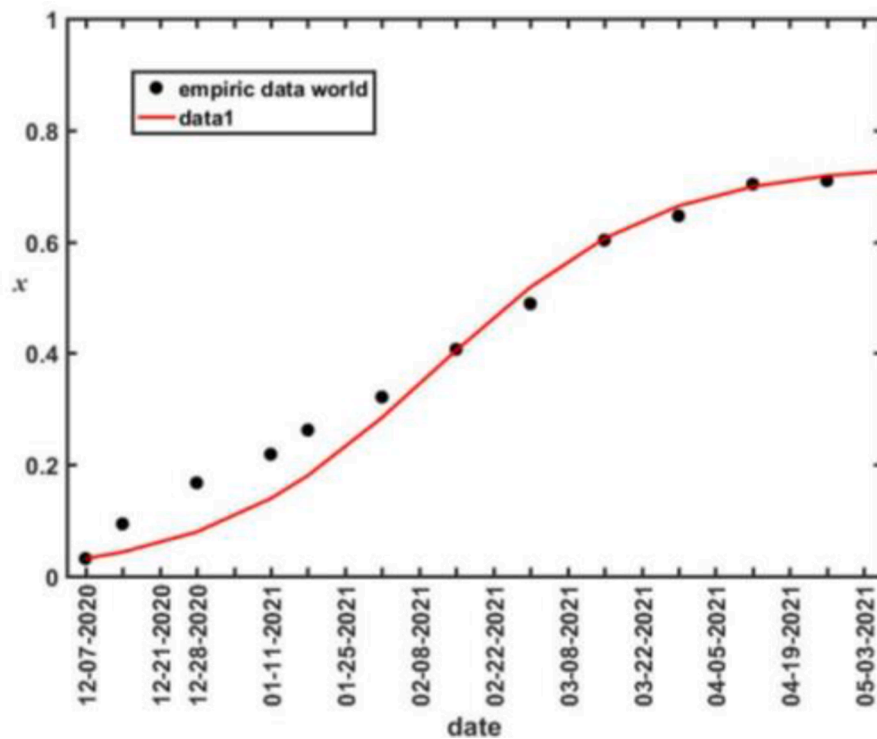

Figure 3. Change in the frequency of B.1.1.7. for the whole world: empirical data (black circles) and formula (4) (solid line).

This parameter can be interpreted as the relative fitness to the local wildtype.

2 The value of this parameter f varies around 1.50 for most of the time series considered in this study. This would imply that the variant is 50% more transmissible than the local wild type.

3 The period in which this variant displaces all or most of the other variants has been in all the studied cases around 22 weeks. This is consistent with the above mentioned almost universal relative fitness f.

4 The model outperforms for cases like Nordjylland which exhibits a very 'jagged' frequency; this often indicates low sequencing numbers
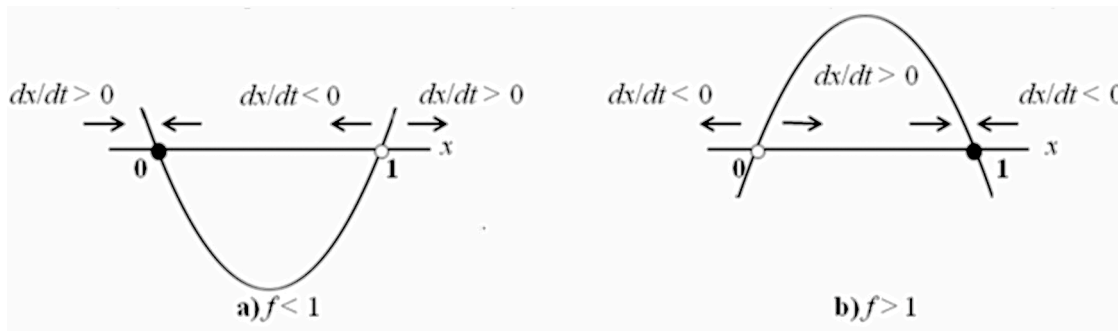
Figure A1. The equilibria of equation (i) and their stability. The arrows alongside the *x*-axis denote the direction of change of *x* when perturbed from equilibrium x^*. If both arrows point towards x^*, then the equilibrium is stable and represented by a filled circle. On the othe hand, if both arrows point away from x^*, then the equilibrium is unstable and represented by an open circle. a) Situation for *f*<1. b) Situation for *f*>1.
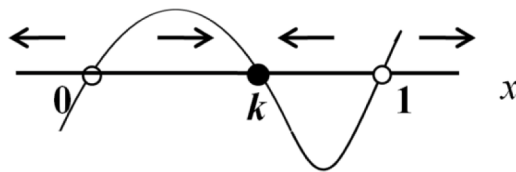
Figure A2. The equilibria of equation (ii) and their stability. Now the stable equilibrium is $x^* = k$.

(Hodcroft 2021) and seems to be the case since this Danish region is the one with lowest statistics of viral sequences.

5 As Hodcroft (2021) stressed, since in many countries sampling may not be equal across the country: samples may only cover one area or certain areas, it's important not to assume frequencies shown are necessarily representative of the country. However, in the case of Denmark it seems there are no important variations across the different regions.

Author contribution

Since I am the only author, I did all the work.

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.

Appendix. Mathematical Modeling: Assumptions and Stability of the Equilibria

Let us enumerate the approximations and assumptions of the modeling used here.

1 All the spatial heterogeneities are neglected, so this is a *mean field* model (Fort 2020).
2 The model distinguishes the variant of concern (VOC), the B.1.1.7 variant, from all the others, which are treated just as one 'mean' lineage. In community and population ecology this is the so-called *focal species* approximation (Fort 2021; Fort 2020).
3 The fitness of the VOC relative to the rest, *f*, is assumed as an intrinsic property of this variant, taking a constant value.
4 In addition, the time frame considered is such that mutations changing *f* can be ignored.

Under these assumptions the Eq. (3) governs the dynamics of the VOC's fraction, *x*. And this equation has two possible equilibria x^* satisfying

$$(f-1)x*(1-x*) = 0. \tag{A1}$$

The stability of both equilibria are shown in Figure A1 below, i.e. $x^* = 0$ (if $f < 1$) and $x^* = 1$ (if $f > 1$).

Since we are taking $f > 1$, the only possible stable equilibrium is $x^* = 1$. This is consistent to what we observe for most of the studied cases.

However, in cases like Switzerland, $x^* = k < 1$. In other words, this implies that there is an additional equilibrium, i.e. instead of (i) we have:

$$(k-x*)x*(1-x*) = 0. \tag{A2}$$

including an extra factor $(k-x^*)$. This slightly more complex and, from an ecological viewpoint, more interesting situation, in which the VOC can coexist with other variants corresponds to the phenomenon of *frequency-dependent selection* (Nowak 2006, chapter 4), that is when the fitness of a phenotype or variant depends on its frequency, $f = f(x)$. In other words, the fitness of variants is determined by the whole 'ecosystem' in which it lives −combinations of coexisting variants and the abiotic environment. Notice that if $f(x)$ is taken proportional to $(x+1-k)$, then eq. (A1) transforms into eq. (A2), and fig. A1-b transforms into fig. A2. Therefore the equilibria $x^* = 1$ became unstable, and the only stable equilibrium is now $x^* = k$.

The simplest way to model this frequency dependent selection is through the well known *replicator equation* (Nowak 2006, chapter 4) given by:

$$\frac{dx}{dt} = x\left[ax + b(1-x) - \overline{f}\right] \tag{A3}$$

Where *a* and *b* are the "payoffs" received by the VOC when interacting with itself and average of all the other variants. Thus it is straightforward to obtain that the VOC "carrying capacity", *k*, is given by:

$$k = \frac{b-1}{b-a}. \tag{A4}$$

References

Latif, A.A. et al., and the Center for Viral Systems Biology. B.1.1.7 Lineage Report 2021. https://outbreak.info/situation-reports?pango=B.1.1.7&loc=GBR&loc=USA&loc=USA_US-CA&selected=Worldwide Accessed 21 June 2021.

Danish Covid-19. 2021. Danish Covid-19 Genome Consortium https://www.covid19genomics.dk/statistics Accessed 21 June 2021.

Davis, N et al. 2020. Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. doi: 10.1101/2020.12.24.20248822 https://www.medrxiv.org/content/10.1101/2020.12.24.20248822v3 Retrieved 30 March 2021.

Fort, H. 2021 Predicting the Yields of Species Occupying a Single Trophic Level with Incomplete Information: Two Approximations Based on the Lotka-Volterra Generalized Equations. bioRxiv 2020.12.31.425009; doi: 10.1101/2020.12.31.425009.

Fort, H., 2020. *Ecological Modelling and Ecophysics: Agricultural and environmental applications* (IOP ebooks). IOP, Bristol, UK, pp. 0–8. Chapter 0.

Hodcroft, E. B. 2021. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. https://covariants.org Accessed 21 June 2021.

ISPM, 2021. Transmission of SARS-CoV-2 variants in Switzerland, 5 March 2021. Retrieved 30 March 2021. https://ispmbern.github.io/covid-19/variants/.

Leung, K., Shum, M.H.H., Leung, G.M, Lam, T.T.Y., Wu, J.T., 2021. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance. Eur. Centre Dis. Control Prevent.* (ECDC) 26 (1), 2002106.

Nowak, M., 2006. Evolutionary Dynamics: Exploring the Equations of Life. Belknap Press, Canada. HOMENEXTSTRAINSTATISTICSABOUT US.