

Inherent random fluctuations in COVID-19 outbreaks may explain rapid growth of new mutated virus variants

Kenneth Bodin^{1†}, Joacim Rocklöv^{2*}

¹High Performance Computing Center North, Umeå University, Umeå, Sweden and Algoryx Simulation, Umeå, Sweden.

²Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden.

*Correspondence to: joacim.rocklov@umu.se

† kenneth.bodin@umu.se

Abstract

A new virus variant of SARS-COV-2 has had a profound impact on society while governments have taken action to limit its impacts by enforcing lockdowns and limiting spread from the UK to other countries. Variants with mutations in the virus genome are likely to occur, but do not always associate to significant changes in the biology of the virus, or the disease. For the variant VOC 202012/01 (also referred to as B.1.1.7), however, preliminary reports indicate it may be more transmissible. Here we use a simulation model calibrated to the inherent random fluctuating transmission pattern of COVID-19 to investigate what the probability may be for detecting more transmissible virus variants post facto. We find that post facto identification of successful virus variants of SARS-COV-2 are likely to exhibit growth rates that are substantially larger than the average growth rate. This finding has implications for interpreting growth rate and transmissibility of new virus variants.

Main text

On December 14, authorities in the United Kingdom reported that a new mutated variant of SARS-CoV-2 had been sequenced, referred to as VOC 202012/01 or B.1.1.7. The new variant could make preventive actions to suppress the COVID-19 pandemic much harder (until effective vaccines have been rolled out), if the preliminary reports of increased transmissibility are correct (1). Therefore, many governments have implemented policies to avoid importing, and to slow the importations of the new virus variant to their country, such as canceling flight traffic, testing and quarantine of air travellers from the UK, border closure for freight transports and trucks, and increased local efforts to survey the circulation and prevalence of virus strains.

Behind triggering the alarm stands both biological and epidemiological insights and data. The SARS-COV-2 variant VOC 202012/01 has 17 mutations, some of potential suspected biological significance linked to its transmissibility (2). Complementary data analysis indicates that this variant could be 56% (2) or 71% (3) more transmissible, and likewise proportionally affect the reproduction number, R. Other reports indicate an increase in R in the range of 0.39-0.93 (3).

Without excluding any biological changes of significance affecting transmissibility, we provide here a potential alternative explanation of the observed patterns of the new virus variants, such as VOC 202012/01, when they are observed dominant in retrospect. Due to the seemingly random inherent fluctuating patterns of the transmission of SARS-CoV-2, we can illustrate using simulations that post-facto discovered rapid growth of new variants may well be explained by the heterogeneous reproduction that characterizes SARS-CoV-2. In fact, we show that specific variants of the virus that survive for many generations are likely to be associated with higher R_e and more dominance due to a majority of variants ending by stochastic extinction. We find that only a smaller proportion of biologically equal variants manage to survive up to 10 generations, while most of them die out early. The difference between survival and extinction can be an early advantageous stochastic event, such as a superspreading.

50 A recent Chinese study with rigorous contact tracing indicates that 15% of all primary infections of
51 SARS-CoV-2 caused 80% of all secondary infections (5). This has also previously been seen in contact
52 tracing data (6), in genome viral sequencing data (7), and in epidemiological data analysis (8). The
53 individual reproductive numbers (i.e. the number of secondary cases originating from one infection)
54 associated with such inhomogeneous spreading patterns can be described well by a heavy tail distribution,
55 typically the negative binomial distribution (5,9). While the individual reproduction number measures the
56 individual inhomogeneous patterns of transmission, the R value measures the average number over all
57 individual secondary infections in a population.

58
59 Inspired by Althouse et.al. (10) we apply a branching model with a negative binomial distribution (NB).
60 We set the effective reproductive number of the NB, $R=1.3$, to represent the current range of the recent R
61 in the UK (1). We use a dispersion parameter of $k=0.30$, taken from Sun et al (5) representing a
62 conservative suppression situation. The generation time is not of large importance to our simulation, but
63 has been estimated to around 5.3 days. However, it can be shorter with effective case isolation and longer
64 with ineffective case isolation (5). From this parameterisation, we grow the transmission chains trajectories
65 iteratively. The index case is one infected individual, which infects a new number of individuals according
66 to a random number generated from the NB distribution. This yields the number of infected in the second
67 generation. For each individual infected in this second generation, we draw a new random number for the
68 third generation, and so on. Henceforth, we establish the full transmission chain from each index case up to
69 generation 10 and we refer to these as individual transmission chain trajectories. Assuming only one virus
70 variant is circulating, these individuals will all be infected by the same virus variant with the same
71 biological characteristics and transmissibility. Repeating this process a large number of times can help us
72 understand what the natural variability of the transmission trajectories are. In fact, the experiment will
73 allow us to estimate how often the R_e value of a trajectory, i.e., virus variant, would be higher than the
74 average R_e value of known virus variants by pure chance. It thus serves as an important natural control to
75 any data driven investigation of detecting important new alternations of R_e , or i.e., transmissibility, of new
76 virus genetic variants.

77
78 We simulate 10 generations, corresponding to approximately 53 days, and estimate transmission chain
79 trajectories for 100,000 infected index individuals carrying the same virus variant. We run the branching
80 processes simulation and track each transmission chain following from each index case for the 10
81 generations. We assume that immunity is already accounted for in R and does not further affect the
82 reproduction during the simulation. Figure 1 illustrates all transmission chain trajectories and the number
83 of individuals being infected from each index case up to generation 10. We observe that many of the
84 transmission chains exhibit stochastic extinction. The average R_e of the surviving trajectories at generation
85 10 is 1.6. This is approximately 23% higher than the actual R of the NB. Using a constant, deterministic
86 number of infections generated in each generation of R_e equal to 1.7, it closes in and equals the average of
87 the surviving trajectories of $R=1.3$, first at generation 8.

88
89 In Figure 2, we show that a high number of strains from the 100,000 index infections die out before the
90 10th generation, in fact, as many as 87.9% go extinct before or at generation 10. This is well known to
91 occur in real life situations, but the frequency of occurrence is related to the stochastic superspreading
92 pattern of SARS-COV-2 and sensitive to the parameter k and the R (11). We find that as many as 20.4% of
93 the surviving strains could be expected to have an R_e of 1.7 or above, substantially higher than the average
94 R_e of 1.3 for all trajectories. On the other hand, a virus variant with even higher transmission, say an R_e of
95 2, would generate 1024 infections in generation 10 of Figure 1, and would be an outlier in the histogram of
96 Figure 2 with a probability as small as 0.024%. We note, however, that the difference to the average
97 surviving trajectories average 1.6 is much less than to the overall average of the NB distribution of 1.3.

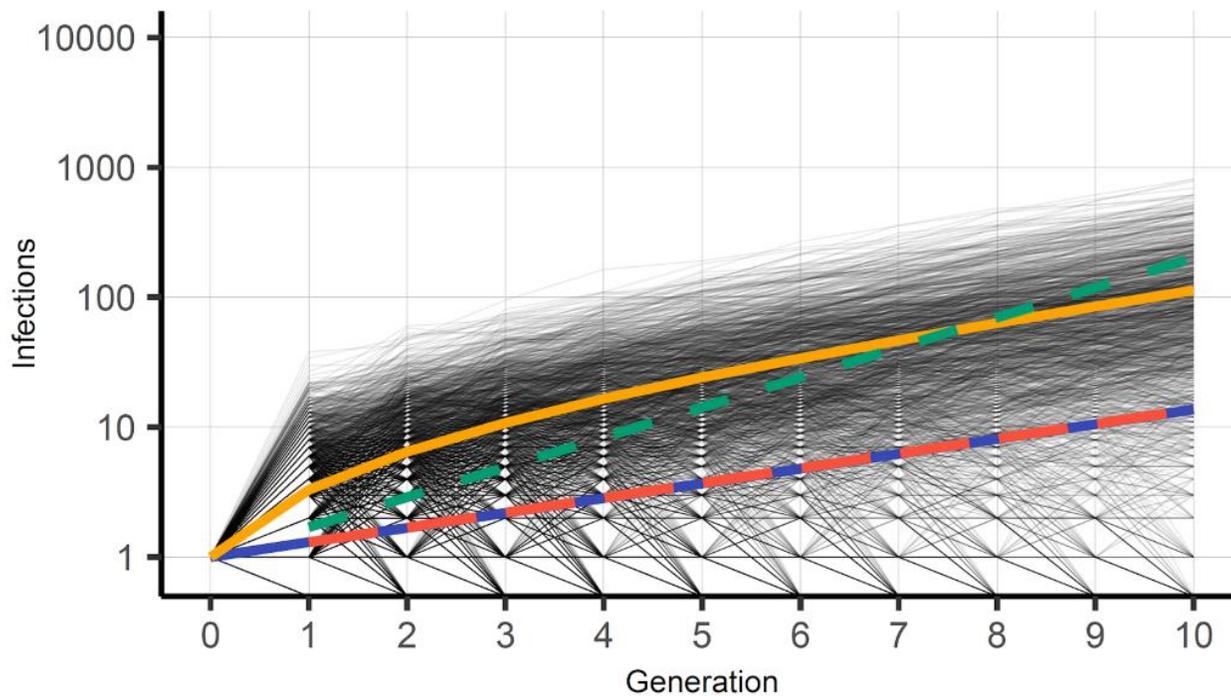
98
99 When we change the parameter k to 0.15, corresponding to a dispersion parameter of a non-suppressed
100 outbreak (7), the extinction rate in generation 10 increases to 92.1% and the probability of a surviving
101 trajectory with R_e greater than 1.7 becomes 34.9%. The probability for a trajectory with R_e greater than
102 2.0 increases to 1.4%. Thus, the effect from random fluctuations is expected to be even more prominent in
103 communities with a lower degree of non-pharmaceutical interventions or lower compliance to regulations,
104 in contrast to China where COVID-19 was effectively suppressed.

106 Our simulation experiment indicates, given that the modelling assumptions would hold in reality, that there
107 is a probability that any post-facto observed virus variant, such as the VOC 202012/01, can in fact be the
108 product of just a very successful random transmission chain trajectory without alternations in the
109 transmissibility, at least this is probable to happen by chance to some degree. Here we show that 20.4% of
110 the transmission trajectories show at least an R_e shift of +0.4 higher than the average. It is important to
111 adjust for this in future studies of changes in transmissibility and R_e values of new virus variants, such as
112 VOC 202012/01. This said, the simulations indicate that the longer the generations are run, pattern
113 stagnates, and this may be less of a problem in situations where the number of generations observed are
114 very large. Finally, we want to stress that we are not providing any empirical proof in any way that the
115 VOC 202012/01 is not more transmissible, it may well be, but we do show that there is a bias when
116 observing post facto successful virus variants, which need to be adjusted for in any data analysis of
117 transmissibility. This has bearing for emerging strains and policy implications for COVID-19, as well as
118 for other infectious diseases, especially those exhibiting substantial inhomogeneous spreading patterns.

119 References

- 121 1. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01. Technical Briefing 2.
122 Public Health England, Dec 28 2020. [https://www.gov.uk/government/publications/investigation-](https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201)
123 [of-novel-sars-cov-2-variant-variant-of-concern-20201201](https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201)
- 124 2. Davies et al. Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern
125 202012/01 in England <https://www.medrxiv.org/content/10.1101/2020.12.24.20248822v1>
- 126 3. NERVTAG, NERVTAG meeting on SARS-CoV-2 variant under investigation, VUI-202012/01, Dec
127 18 2020.
128 [https://khub.net/documents/135939561/338928724/SARS-CoV-](https://khub.net/documents/135939561/338928724/SARS-CoV-2+variant+under+investigation%2C+meeting+minutes.pdf/962e866b-161f-2fd5-1030-32b6ab467896)
129 [2+variant+under+investigation%2C+meeting+minutes.pdf/962e866b-161f-2fd5-1030-32b6ab467896](https://khub.net/documents/135939561/338928724/SARS-CoV-2+variant+under+investigation%2C+meeting+minutes.pdf/962e866b-161f-2fd5-1030-32b6ab467896)
- 130 4. Coronavirus (COVID-19) Infection Survey, UK Statistical bulletins,
131 [https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseas](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases)
132 [es/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases)
- 133 5. K. Sun, W. Wang, L. Gao, Y. Wang, K. Luo, L. Ren, Z. Zhan, X. Chen, S. Zhao, Y., Huang, Q.
134 Sun, Z. Liu, M. Litvinova, A. Vespignani, M. Ajelli, C. Viboud, H. Yu, Transmission
135 heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* 10.1126/science.abe2424
136 (2020). doi:10.1126/science.abe2424
137 <https://science.sciencemag.org/content/early/2020/11/23/science.abe2424>
- 138 6. Adam, D.C., Wu, P., Wong, J.Y. et al. Clustering and superspreading potential of SARS-CoV-2
139 infections in Hong Kong. *Nat Med* 26, 1714–1719 (2020). [https://doi.org/10.1038/s41591-020-](https://doi.org/10.1038/s41591-020-1092-0)
140 [1092-0](https://doi.org/10.1038/s41591-020-1092-0)
- 141 7. Miller, D., Martin, M.A., Harel, N. et al. Full genome viral sequences inform patterns of SARS-
142 CoV-2 spread into and within Israel. *Nat Commun* 11, 5518 (2020).
143 <https://doi.org/10.1038/s41467-020-19248-0>
- 144 8. Akira Endo, Sam Abbott, Adam J Kucharski, Sebastian Funk. Estimating the overdispersion in
145 COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res*, **5**, 67 (2020).
146 <https://doi.org/10.12688/wellcomeopenres.15842.3>
- 147 9. J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, W.M. Getz. Superspreading and the effect of
148 individual variation on disease emergence. *Nature*. 2005; 438:355–359.
149 <https://doi.org/10.1038/nature04153>
- 150 10. Benjamin M. Althouse, Edward A. Wenger, Joel C. Miller, Samuel V. Scarpino, Antoine Allard,
151 Laurent Hébert-Dufresne, Hao Hu. Superspreading events in the transmission dynamics of SARS-
152 CoV-2: Opportunities for interventions and control. *PLOS Biology*, (2020)
153 <https://doi.org/10.1371/journal.pbio.3000897>
- 154 11. Hartfield, M., & Alizon, S. (2013). Introducing the outbreak threshold in epidemiology. *PLoS*
155 *pathogens*, 9(6), e1003277. <https://doi.org/10.1371/journal.ppat.1003277>

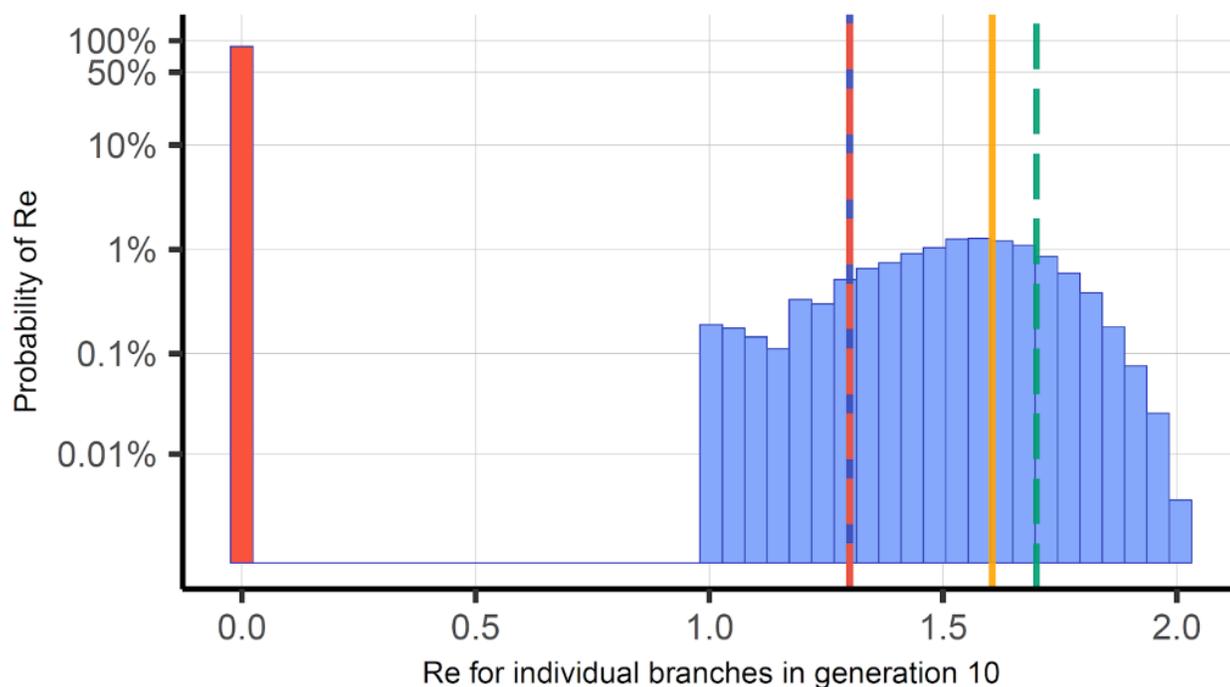
159
160



161
162
163
164
165
166
167
168

Figure 1. The trajectories of new infections at each generation resulting from the 100,000 index infections and biologically similar strains at generation 0 (grey lines). The yellow line shows the average number of infections among the trajectories that are sustained (only 12.1% in generation 10). The blue line shows the deterministic number of infections from R equal to 1.3 and the dashed red line shows the corresponding average measured in the simulation. The green dashed line shows the deterministic number of infections if R is equal to 1.7, which crosses the average of surviving trajectories of Re 1.3 first at generation 8.

169



170
171

172 **Figure 2.** The frequency of the R_e 's of the different transmission chain trajectories after 10 generations.
173 The strains that go extinct are described by the red bar and equals 87.9%. The remaining R_e estimates fall
174 between 1.0 to 2.0. R_e of 1.3 is marked by the red and blue vertical line, the average R_e of the surviving
175 transmission chains is 1.6 and marked by the vertical yellow line. 20.4% of the surviving transmission
176 chains have R_e values above 1.7 by random (dashed green line).

177
178
179

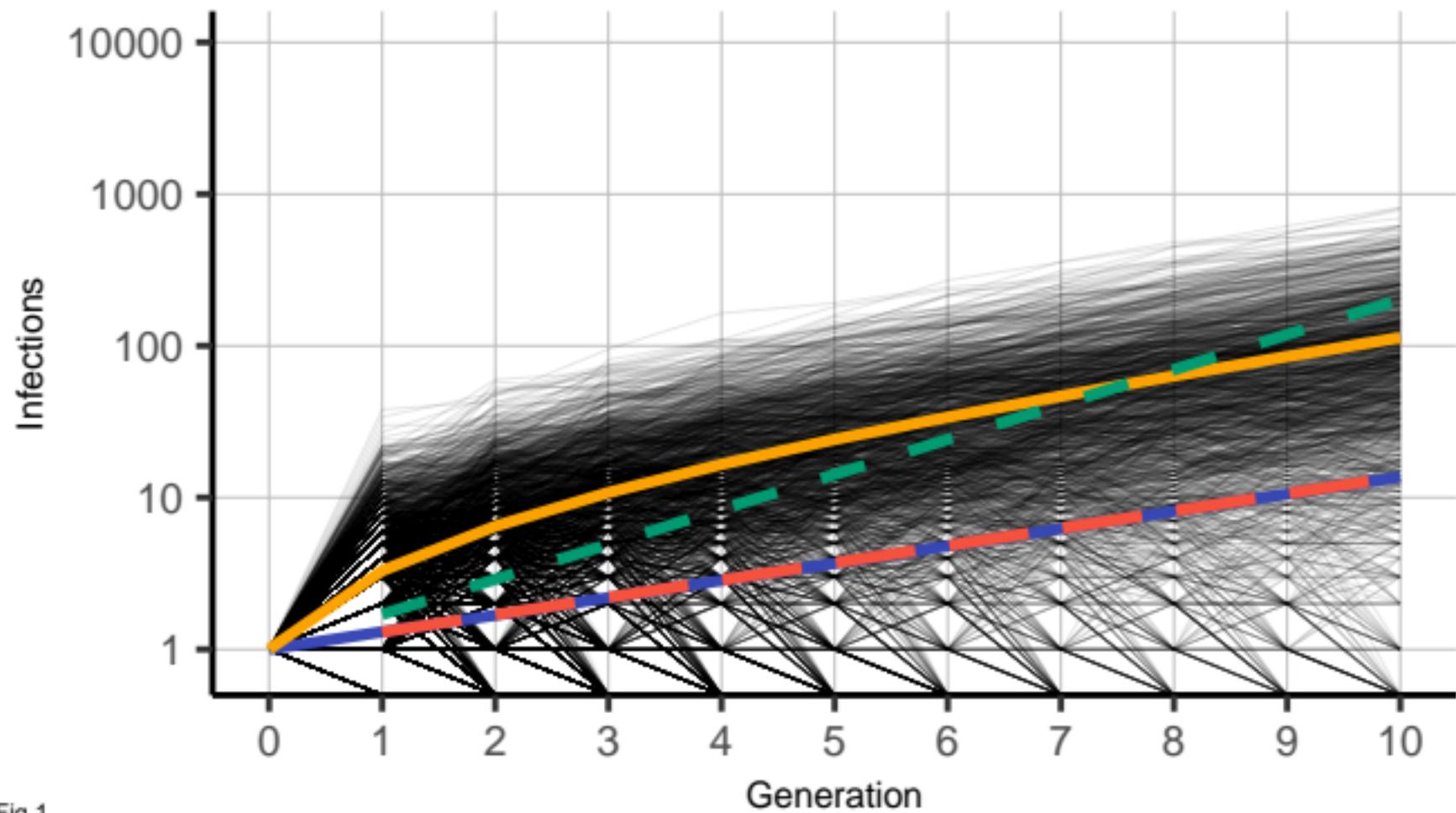


Fig 1

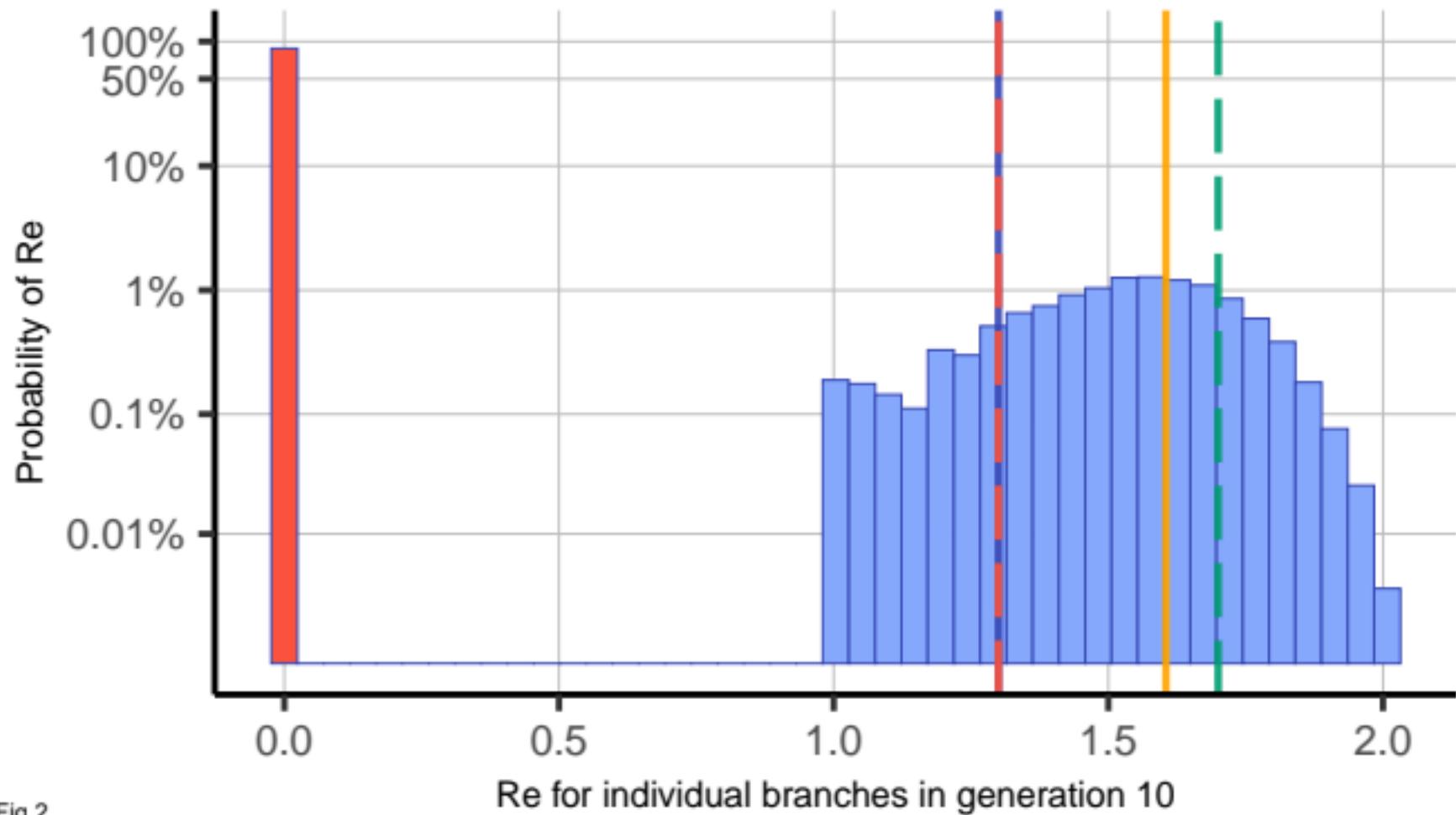


Fig 2