



# Selective sweeps in SARS-CoV-2 variant competition

Laura Boyle<sup>a</sup>, Sofia Hletko<sup>a</sup>, Jenny Huang<sup>a</sup>, June Lee<sup>a</sup>, Gaurav Pallod<sup>a</sup>, Hwai-Ray Tung<sup>a</sup>, and Richard Durrett<sup>a,1</sup>

Contributed by Richard Durrett; received August 12, 2022; accepted September 15, 2022; reviewed by Jasmine Foo and Natalia Komarova

The main mathematical result in this paper is that change of variables in the ordinary differential equation (ODE) for the competition of two infections in a Susceptible–Infected–Removed (SIR) model shows that the fraction of cases due to the new variant satisfies the logistic differential equation, which models selective sweeps. Fitting the logistic to data from the Global Initiative on Sharing All Influenza Data (GISAID) shows that this correctly predicts the rapid turnover from one dominant variant to another. In addition, our fitting gives sensible estimates of the increase in infectivity. These arguments are applicable to any epidemic modeled by SIR equations.

covid variant competition | selective sweep | logistic equation

As most readers know, the COVID-19 pandemic began in China in December 2019, then slowly spread around the world. By early 2022, there had been more than 400 million infections worldwide and almost 6 million deaths. The first confirmed case of COVID-19 in the United States was diagnosed in Washington state on January 21, 2020. On February 11, 2020, the disease was officially named “severe acute respiratory syndrome coronavirus 2” (SARS-CoV-2) because the virus is genetically related to the coronavirus responsible for the SARS outbreak of 2003. The early spread of the virus is an interesting topic. Work done by Alessandro Vespignani’s group (1) suggested that on March 1, 2020, when New York recorded its first case and there were only 23 confirmed cases in the United States, there could have actually been about 28,000 infections nationwide and 10,700 in New York. Many papers have been written on the effects of interventions such as masking, lockdowns, and social distancing to control the spread of the disease. Here, we will concentrate on the evolution of the virus, primarily focusing on the changes in the spike protein, which the virus uses to gain entry into cells through binding to the ACE2 surface protein. (For a detailed description of this process, see refs. 2 and 3.) Figure 1 shows the rise and fall of variants over time. Figure 2 indicates their phylogenetic relationship.

## Early Evolution

The article by Koelle et al. (4) gives a nice account of the evolution of the virus up to the end of February 2022. An early mutation in the spike protein was a glycine residue, G, replacing an aspartic acid residue, D, at position 614, or G614D for short. (There are 20 amino acids, each abbreviated to a single letter, but the exact coding is not important for our purposes.) This dramatically increased the ability of the original Wuhan strain to infect cells.

The Alpha variant (named B.1.1.7 for its position in the phylogenetic tree of samples) was identified in the United Kingdom in September 2020. It had enhanced binding to ACE2 and spread 50% faster than circulating lineages. Around the same time, Beta (B.1.351) was found in South Africa, and a highly transmissible variant, Gamma, arose in the Amazonas state in Brazil.

The Delta variant was identified in India’s Maharashtra state in Spring 2021. It had additional mutations that increased the ability of the spike protein to infect cells, resulting in increased transmissibility, disease severity, and breakthrough infections in vaccinated individuals (5, 6). A notable example is the P681R mutation, located at a furin cleavage site that separates the spike protein into S1 and S2 subunits (7). A second is the L452R mutation in the receptor binding domain of the spike protein (8), which is involved in evasion from neutralizing antibodies. Once Delta arrived in the United Kingdom, it spread quickly, and epidemiologists determined that it was about 60% more transmissible than Alpha. In vitro, it is 6 times less sensitive to serum-neutralizing antibodies from recovered individuals compared to the wild-type Wuhan-1 bearing D614G. It swept through India and Great Britain before reaching the United States, where it surged from being 1.3% of variants on May 2, 2021, to 94.4% on July 31.

## Omicron Subvariants

Two articles by Ewen Callaway (9, 10) give an excellent account of this phase of the epidemic. Omicron (B.1.1.529) was first reported in South Africa in November 2021.

## Significance

The main contribution of this paper is to show that the transition from one SARS-CoV-2 variant to the next is mathematically equivalent to a selective sweep, which follows the logistic equation. Data show that this prediction is accurate. The increase in fitness from one variant to the next has two components: the increase in transmissibility and the contribution of breakthrough infections. This predicts that future dominant strains will be the ones that best evade immunity.

Author affiliation: <sup>a</sup>Department of Mathematics, Duke University, Durham, NC 27708-0320

Author contributions: R.D. designed research; L.B., S.H., J.H., J.L., G.P., and H.-R.T. performed research; and J.H., H.-R.T., and R.D. wrote the paper.

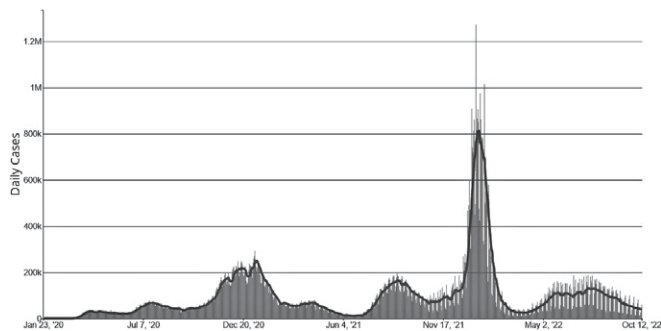
Reviewers: J.F., University of Minnesota; and N.K., University of California, Irvine.

The authors declare no competing interest.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: rtd@math.duke.edu.

Published November 3, 2022.

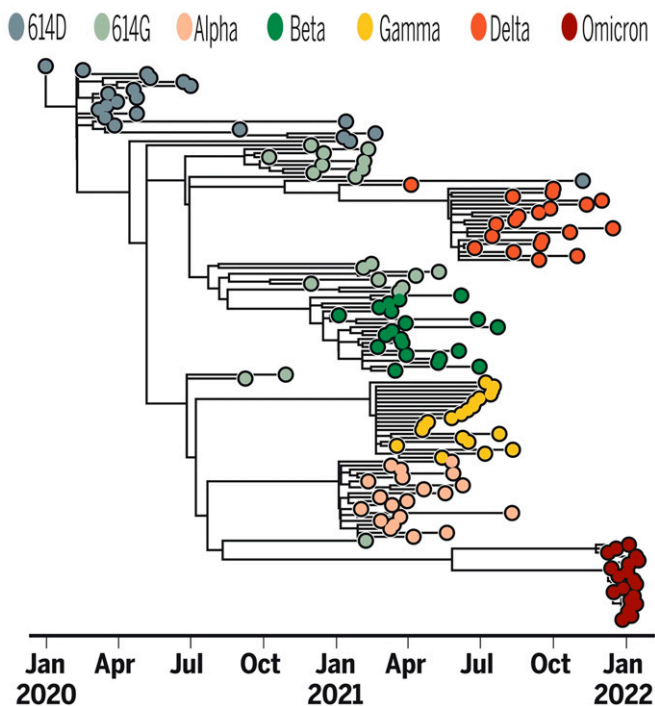


**Fig. 1.** Epidemic waves in the SARS-CoV-2 pandemic. From left to right: Beta, Delta, early Omicron variants, and Omicron BA.4/BA.5. Graphic is from *The New York Times*, July 18, 2022.

It has 37 mutations in the spike protein relative to the original Wuhan strain. Among these mutations, 15 were in the receptor binding domain, increasing its binding efficiency to ACE2 (11). An early article (12) suggested that the most recent common ancestor of Delta and Omicron went back to the beginning of the pandemic, but the picture changed when Omicron was realized to be a close relative of Gamma.

After a period of using Greek letters to mark significant changes in the virus, the system broke down when a large number of Omicron variants started appearing. See ref. 13 for the story behind the names. The initial variant, now referred to as BA.1, has continued to evolve, creating two sublineages: BA.1.1 = BA.1 with an R346K mutation and BA.2. The BA.2 subvariant spreads more quickly than BA.1. Two laboratory studies have shown that antibodies against BA.1 can protect against BA.2 infection, but they have similar abilities to resist neutralizing antibodies in the blood of individuals that have been vaccinated or infected (10, 14). For the reader who is interested in the mutations found in the early Omicron variants, see ref. 15 for an analysis of 6.4 million sequenced genomes.

Studies by virologists Michael Chan at the University of Hong Kong and Wendy Barclay at Imperial College London have shown



**Fig. 2.** SARS-CoV-2 variant phylogeny from ref. 4.

that BA.1 and BA.2 create a higher concentration of viral particles in the nose. In earlier variants, after SARS-CoV-2 binds to ACE2, it relies on a cellular enzyme, TMPRSS2, to cleave the spike protein, granting the virus entry into the cell, but Omicron has largely abandoned this route. This could give Omicron an edge, since many cells in the nose do not make this enzyme, which is more widely expressed in the lung and other organs (16).

### Omicron Subvariants

**BA.4 and BA.5.** The earliest samples of variants BA.4 and BA.5 were collected in January 2022 in South Africa. Generally, they cause mild disease, but spread in large numbers potentially because, unlike the Wuhan strain, which settles in the lungs, these new variants seem to attach to the more benign upper nasal passage. The incubation time of these new variants is significantly shorter (2 to 3 d), and they seem to undergo mutational sprints, mutating as much as 4 times faster than normal Omicron. Like earlier Omicron variants, they have a remarkable ability to evade immunity from vaccines, previous infection, or both (17–19). Figure 3 shows the rise and fall of Omicron variants over time.

### Results

We use a traditional Susceptible–Infected–Removed (SIR) epidemic model in a homogeneously mixing population. In many models, 1) an Exposed phase is included, in which individuals have the disease, but are not yet infectious; and 2) the population is divided into a half-dozen groups according to age (20–22). However, we choose the simplicity of the SIR model in order to easily make clear mathematical statements that reveal the dependence of the observed phenomena on underlying parameters.

To get rid of the population size from the SIR equations, we rewrite them in terms of the scaled variables  $s = S/N$ ,  $i_j = I_j/N$ , and  $r = R/N$ . Generalizing the basic model to the competition of two infections, we can write the differential equations as

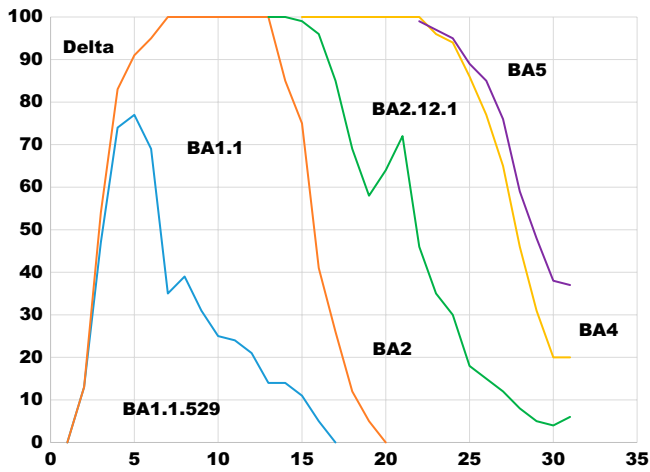
$$\begin{aligned} \frac{ds}{dt} &= -\beta_1 s i_1 - \beta_2 s i_2 \\ \frac{di_1}{dt} &= \beta_1 s i_1 - \gamma i_1 \\ \frac{di_2}{dt} &= \beta_2 (s + \theta r_1) i_2 - \gamma i_2 \\ \frac{dr_1}{dt} &= \gamma i_1 - \theta r_1 i_2 \\ \frac{dr_2}{dt} &= \gamma i_2. \end{aligned} \quad [1]$$

Here, infecteds of type  $i$  at rate  $\beta_i$  attempt to infect a randomly chosen individual from the population. Individuals recovered from infection by strain  $i$  are immune to further infection by that strain.  $\theta$  is the reduction in the infection rate by strain 2 of an individual immune to strain 1.

In *Materials and Methods*, we will show that the fraction of individuals infected with strain 2,  $x(t) = i_2(t)/(i_1(t) + i_2(t))$ , satisfies a logistic differential equation,

$$\begin{aligned} x'(t) &= \lambda(t)x(t)(1 - x(t)) \\ \lambda &= (\beta_2 - \beta_1)s(t) + \theta\beta_2 r_1(t). \end{aligned} \quad [2]$$

When  $\lambda(t)$  is constant, this models a selective sweep, in which an advantageous mutation increases in frequency and eventually takes over the population (23). To simplify the formulas, we assume that the recovery rates are equal for the two strains. If they are different, then  $\gamma_1 - \gamma_2$  is added to the formula for  $\lambda(t)$ .



**Fig. 3.** Weekly Omicron variant frequencies in North Carolina during June 26 through July 9, 2022. Data are from the North Carolina COVID-19 Dashboard.

To test this prediction, we used SARS-CoV-2 variant data from the Global Initiative on Sharing All Influenza Data (GISAID) to fit our model (24). These data consisted of biweekly SARS-CoV-2 variant cases in the United States from the period between December 12, 2021, and June 13, 2022. For each transition—Beta to Delta, Delta to Omicron, and Omicron to BA.4/BA.5—we fit the logistic differential equation to data points at the beginning and end of each selective sweep curve (Table 1).

The increase in infectivity in the Delta to Omicron transition is much larger than Beta to Delta due to breakthrough infections. The increase from Omicron to BA.4/BA.5 is the smallest, since the Omicron subvariants are more similar than Beta and Delta. Another contributing factor is that the size of the susceptible population was decreased by the first Omicron wave.

In fitting the data, we have assumed that  $\lambda(t)$  is constant. The selective sweeps pictured in the top three panels of Fig. 4 took from 6 to 10 wk, so the values of  $s(t)/s(0)$  and  $r_1(t)/r_1(0)$  should not have changed by much over the course of the sweep, but perhaps this is the source of the departure from linearity seen in the bottom panel.

## Discussion

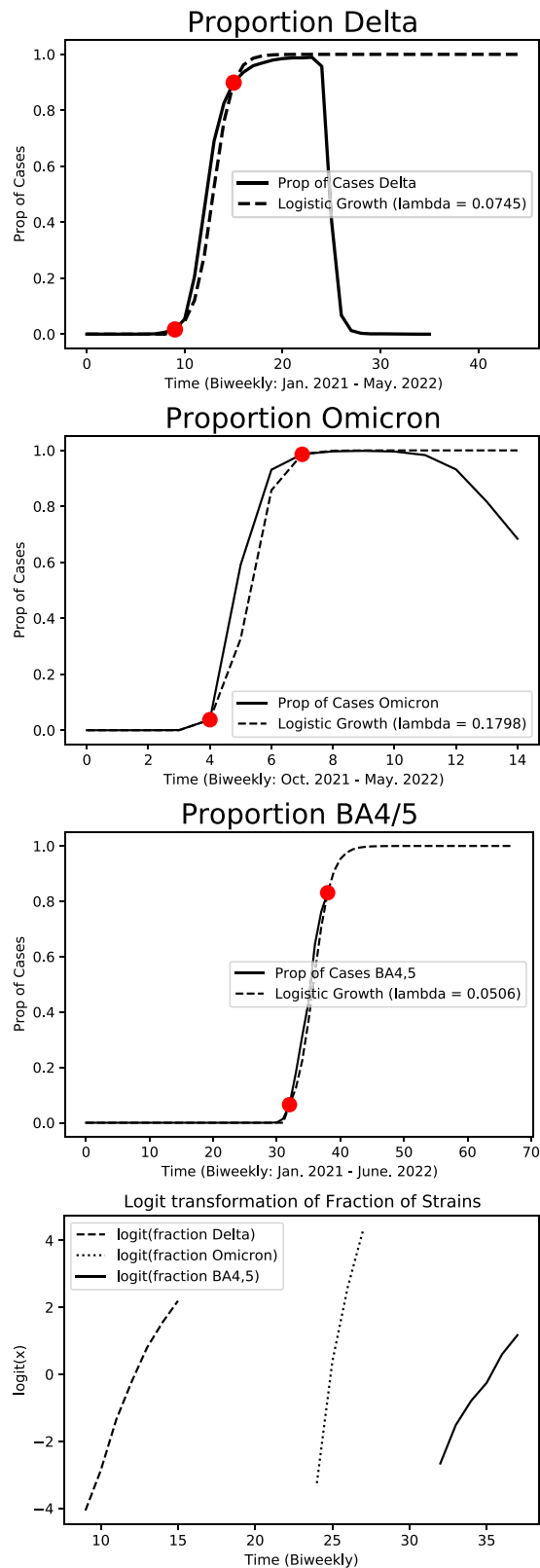
Our results in Fig. 4 show that the rapid turnover from one variant to another follows the solution to the logistic differential equation. The formula for the fitness advantage  $\lambda$  given in Eq. 2 has two terms:

- $(\beta_2 - \beta_1)s(t)$  is the increased infection rate of strain 2;
- $\theta\beta_2r_1(t)$  is the contribution of breakthrough infections.

There have been a number of mutations that have significantly improved the ability of SARS-CoV-2 to infect humans. It is natural to expect that, as time goes on, such gains will become even smaller, since the spike protein has explored the space of

**Table 1. Estimates of  $\lambda$**

Transition	$\lambda$
Beta to Delta	0.0745
Delta to Omicron	0.1798
Omicron to BA.4/BA.5	0.0506



**Fig. 4.** The top three panels show logistic fits to COVID variant transitions. If the fraction of cases due to the new variant,  $x(t)$ , was the solution of a logistic equation, then  $\log(x(t)/(1 - x(t)))$  would be linear. In the bottom panel, we plot this transformation of the three curves. The resulting plots are approximately linear with  $R^2$  values 0.98, 0.97, and 0.98, respectively.

possibilities. Thus, given our formula for  $\lambda(t)$ , strains that become dominant will need to evade existing immunity. Experts agree with this conclusion (9). Sarah Cobey, an evolutionary biologist

at the University of Chicago, stated: “As gains in infectivity start to slow, the virus will have to maintain its fitness by overcoming immunity.” Kristian Andersen of Scripps Research stated: “Variants such as Omicron that gain much of their transmission advantage from evading immune response may become the norm, as is the case for seasonal influenza.”

Given the similarities (and differences) with the flu, it is an important problem to understand the mechanisms of immune evasion in SARS-CoV-2 in order to predict its future evolution. Readers who want to tackle this problem should look at the excellent work of Dushoff, Levin, and Plotkin (25, 26) on influenza.

## Materials and Methods

Here, we derive the result given in Eq. 2. Let  $x(t) = i_2(t)/(i_1(t) + i_2(t))$ . To prepare for the change of variables, we note that

$$\left(\frac{i_2}{i_1 + i_2}\right)' = \frac{i_2'(i_1 + i_2) - i_2(i_1' + i_2')}{(i_1 + i_2)^2} = \frac{i_2'i_1 - i_2i_1'}{(i_1 + i_2)^2}.$$

1. J. T. Davis *et al.*, Estimating the establishment of local transmission and the cryptic phase of the COVID-19 pandemic in the US. *medRxiv* [Preprint] (2020). <https://www.medrxiv.org/content/10.1101/2020.07.06.20140285v2>.
2. J. Shang *et al.*, Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11727–11734 (2020).
3. C. B. Jackson, M. Farzan, B. Chen, H. Choe, Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **23**, 3–20 (2022).
4. K. Koelle, M. A. Martin, R. Antia, B. Lopman, N. E. Dean, The changing epidemiology of SARS-CoV-2. *Science* **375**, 1116–1121 (2022).
5. P. Elliot *et al.*, Exponential growth, high prevalence of SARS-CoV2, and vaccine effectiveness associated with the Delta variant. *Science* **374**, 1463 (2021).
6. P. Mlcochova *et al.*; Indian SARS-CoV-2 Genomics Consortium (INSACOG); Genotype to Phenotype Japan (G2P-Japan) Consortium; CITIID-NIHR BioResource COVID-19 Collaboration, SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**, 114–119 (2021).
7. I. P. Peacock, *The SARS-CoV-2 variants associated with infections in India, B.1.617, show enhanced spike cleavage by furin*. *BioRxiv* (2021). <https://www.biorxiv.org/content/10.1101/2021.05.28.446163v1>.
8. R. K. Gupta, E. J. Topol, COVID-19 vaccine breakthrough infections. *Science* **374**, 1561–1562 (2021).
9. E. Callaway, Beyond Omicron: What's next for COVID's viral evolution. *Nature* **600**, 204–207 (2021).
10. E. Callaway, Omicron sub-variant: What scientists know so far. *Nature* **602**, 556–557 (2022).
11. Q. Hong *et al.*, Molecular basis of receptor binding and antibody neutralization of Omicron. *Nature* **604**, 546–552 (2022).
12. K. Kupferschmidt, Startling new variant raises urgent questions. *Science* **374**, 1178–1180 (2021).
13. A. Maxmen, Why call it BA.2.12.1? A guide to the tangled Omicron family. *Nature* **606**, 446–447 (2022).
14. S. Ikhtani *et al.*, Antibody evasion properties of SARS-CoV-2 Omicron sublineages. *Nature* **604**, 553–556 (2022).
15. F. Obermeyer *et al.*, Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. *Science* **376**, 1327–1332 (2022).
16. A. Dance, Omicron's lasting mysteries: Four questions scientists are racing to answer. *Nature* **603**, 22–24 (2022).
17. G. Vogel, New subvariants are masters of immune evasion. *Science* **376**, 679–680 (2022).
18. P. Katona, Omicron BA.4 and BA.5: Starting from scratch yet again. *Medpage Today* (2022). <https://www.medpagetoday.com/opinion/second-opinions/99082>.
19. L. Leatherby, What the BA.5 subvariant could mean for the United States. *The New York Times*, 7 July 2022. <https://www.nytimes.com/interactive/2022/07/07/us/ba5-covid-omicron-subvariant.html>.
20. T. Britton, F. Ball, P. Trapman, A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**, 846–849 (2020).
21. S. Fox *et al.*, The impact of asymptomatic COVID-19 infections on future pandemic waves. *medRxiv* [Preprint] (2020). <https://www.medrxiv.org/content/10.1101/2020.06.22.20137489v1>.
22. K. Pierce *et al.*, COVID-19 healthcare demand projections: Beaumont-Port Arthus MSA, Texas. [Preprint] (2020). [www.bio.utexas.edu/research/meyers/publications.html](http://www.bio.utexas.edu/research/meyers/publications.html).
23. R. Durrett, J. Schweinsberg, Approximating selective sweeps. *Theor. Popul. Biol.* **66**, 129–138 (2004).
24. B. E. Hodcroft, CoVariants: SARS-CoV-2 mutations and variants of interest (2021). <https://covariants.org/>. Accessed 21 July 2022.
25. J. B. Plotkin, J. Dushoff, S. A. Levin, Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6263–6268 (2002).
26. S. A. Levin, J. Dushoff, J. B. Plotkin, Evolution and persistence of influenza A and other diseases. *Math. Biosci.* **188**, 17–28 (2004).

Using this in Eq. 1, we get

$$\begin{aligned} x'(t) &= \frac{i_2}{i_1 + i_2} \cdot \frac{i_1}{i_1 + i_2} \cdot \left(\frac{i_2'}{i_2} - \frac{i_1'}{i_1}\right) \\ &= \frac{i_2}{i_1 + i_2} \cdot \frac{i_1}{i_1 + i_2} \cdot [(\beta_2 - \beta_1)s + \theta r_1(t)], \end{aligned}$$

writing  $\lambda(t) = (\beta_2 - \beta_1)s(t) + \theta\beta_2 r_1(t)$ . This is a logistic differential equation, with carrying capacity  $K = 1$  and a temporally varying growth rate:

$$x'(t) = \lambda(t)x(t)(1 - x(t)).$$

**Data Availability.** Previously published data were used for this work (GISAID database) (<https://gisaid.org/>).

**ACKNOWLEDGMENTS.** This work was done during the Duke University DO-Math summer research program for undergraduates. R.D., H.R.-T., L.B., S.H., and J.H. were partially supported by NSF Grant DMS 1809967 from the Probability Program. We thank Ivana Bozic, Tom Britton, Jasmine Foo, Natalia Komarova, Kevin Leder, Simon Levin, ShiShi Luo, and Jason Schweinsberg for comments on previous versions that helped improve the paper. They are, of course, not responsible for any problems that remain.