Empirical tools for studying genetic drift in microbial populations

by

QinQin Yu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Oskar Hallatschek, Chair
Professor Hernan G. Garcia
Professor Yun S. Song
Professor Matthew F. Traxler

Spring 2022

Empirical tools for studying genetic drift in microbial populations

Abstract

Empirical tools for studying genetic drift in microbial populations

by

QinQin Yu

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Oskar Hallatschek, Chair

Deciphering the processes that govern microbial evolution allows us to make predictions of systems ranging from pathogen evolution to climate-driven ecosystem shifts. One fundamental evolutionary process is genetic drift, which is the stochastic change in population composition due to the randomness of birth and death processes. Genetic drift can lead to the loss of genetic diversity and weaken the efficacy of natural selection; thus, inferring its strength and understanding how it arises is important for understanding evolutionary dynamics. Despite decades of research on genetic drift, we still have limited empirical tools to infer the strength of genetic drift from data. Additionally, the strength of genetic drift is often considered to be a static property of a population on short timescales rather than an evolvable trait. In this thesis, we develop new methods for empirically measuring the strength of genetic drift and test hypotheses about the mechanisms that set the strength of genetic drift.

In Chapter 2, we test the hypothesis that mutations can change the strength of genetic drift, the first requirement needed for a trait to be evolvable. We focus on microbial colonies, which are a model system for range expansions. To test this hypothesis, we develop a new experimental method to measure the strength of genetic drift in high throughput using fluorescence microscopy. We find that mutations significantly affect the strength of genetic drift by causing changes in the self-organized spatial structure of the colony. These changes to genetic drift substantially affect the probability that new beneficial mutations escape stochastic extinction, providing evidence that genetic drift may be an evolvable property of a population.

In Chapter 3, we investigate the strength of genetic drift in SARS-CoV-2 evolution at the host population level. We develop a statistical inference framework for inferring the strength of genetic drift simultaneously with measurement noise from lineage frequency time series data. Applying this method to genomic data from England, we find that the strength of genetic drift is consistently, throughout time, higher than expected given the number of people

infected with COVID-19 at the community level, even after correcting for measurement noise and epidemiological dynamics. We also find evidence for spatial structure in SARS-CoV-2 transmission at the regional level. The levels of genetic drift that we observe are higher than the estimated levels of superspreading found by modeling studies that incorporate data of contact statistics in England. We discuss how even in the absence of superspreading, high levels of genetic drift can be generated via jackpot events in a deme model.

The new experimental and computational methods developed in this thesis allow us to infer the strength of genetic drift, and thus gain a better understanding of evolutionary dynamics, in a larger range of laboratory and natural settings. We find that random single mutations can change the strength of genetic drift and affect downstream evolution, suggesting that genetic drift can be an evolvable trait of a population. Finally, we find that multiple mechanisms can alter the strength of genetic drift at the population level.

To my family.

Contents

Acknowledgments

I am deeply grateful to my advisor, Oskar Hallatschek, for teaching me theoretical tools, asking important questions, shaping my scientific interests and approaches, and for supporting me in various scientific and non-scientific endeavors. To Marie Cécilia-Duvernoy, Matti Gralka, Jona Kayser, Benjamin Good, and Alison Feder: thank you for teaching me how to do research; for teaching me lab, analysis, and communication skills; for supporting me when the work was tough; and for motivating and inspiring me to become a better scientist. I am also grateful to my thesis committee members, Hernan Garcia, Matt Traxler, and Yun Song, for feedback on my projects and advice about navigating a scientific career.

I was fortunate to have had the opportunity to collaborate and learn from Joao Ascensao, Aditya Prasad, Arbel Harpak, Megan Sousa, Takashi Okada, Annie Cheng, Can Goksal, and Kristen Lok. I am particularly grateful to Joao for teaching me molecular biology, statistical inference, and microbiology, and for teaching me by example how to be a rigorous and motivated scientist. I am deeply grateful to the entire Hallatschek group for being encouraging and kind colleagues, and for making doing and discussing science together so fun.

I am grateful to the many scientific communities that have generously allowed me to join them and have made my research endeavors intellectually exciting and fun: the Center for Theoretical and Evolutionary Genomics, Ecology and Evolution of Infectious Diseases, QB3 (Quantitative Biosciences), and the Plant and Microbial Biology Department. The Physics Department has been my home and I feel lucky to have interacted with many brilliant and kind scientists there, particularly Naomi Ginsberg and Holger Mueller, who have inspired me by their research and mentorship. I am grateful to the entire Martin lab, particularly Ken Dong and Ellen Goodall, for answering many questions about lab techniques, allowing use of their equipment, lending reagents, and creating a positive lab atmosphere. The MBL Physiology course was a formative experience for me, and all of the instructors and students inspired me by their curiosity, excitement, and persistence, and importantly, they also taught me the fundamentals of research in biology! I hope to continue embodying the Woods Hole spirit throughout my career.

The administrators and staff of the Physics Department and QB3 have helped me take care of countless administrative matters and have supported me in many more ways. I would particularly like to acknowledge Joelle Miles and Anne Takizawa in the Physics Department and Arezou Razavi and Claris Garzon in QB3 for making things happen smoothly behind the scenes and for answering my countless questions.

Thank you to the many friends who have made my time in grad school so fun and worthwhile, for lending a listening ear when things got tough, and for helping me grow as a person. Marie Lu, Ran Fan, Mengfei Sun, Bryan Au-Yeung, Jonathan Liu, Peter Madigan, Andrea Egan, Emma Stansbury, Conrad Stansbury, and Pra Bhattacharyya, thanks for all the dinner parties, picnics, game nights, walks, birthdays, and more. Thank you to the Science Policy Group and the awesome ladies of my string quartet, Jacquelyn Ho, Kristen

Cotner, Claire Gasque, and Indu Chandrasekhar, for helping me discover and sustain my interests outside of research.

I have been shaped into the scientist and person I am today by all of my mentors and teachers, I would particularly like to acknowledge my high school science teachers and my undergraduate professors, Anna Frebel, David Pritchard, and Vladan Vuletic for teaching me what it means to do science and giving me opportunities to gain experience early on in my journey in science.

I am deeply grateful to my earliest supporters and mentors, my parents Lixin Ma and Ping Yu; thank you for exposing me to the joys of science and research, and your endless love and support throughout the years. And last but not least, thank you to my partner in life, Jakob Dahl, for sharing in this life chapter together, learning together, your endless love and support, and bringing much joy to my life.

Chapter 1

Introduction

Microbial evolution is ubiquitous in life, and occurs in settings ranging from infectious disease to food production and environmental health. Deciphering the processes that govern microbial evolution allows us to understand how microbial evolution occurs in these different systems, test hypotheses for why it occurs the way it does, and ultimately perhaps even to shape microbial evolution to benefit environmental and human health.

In order to understand microbial evolution, we have to understand the different processes that shape it: mutation, recombination, natural selection, genetic drift, and migration [45]. These different processes create and remove genetic variation, changing the abundance of different genotypes (the genetic makeup of individuals) in the population over time. Mutation and recombination create genetic variation: mutation is the stochastic change in genotype from parent to offspring due to replication error, and recombination is the creation of a new genotype by combining two existing genotypes. On the other hand, natural selection and genetic drift reduce genetic variation: natural selection is the preferential replication of particular genotypes that are most suited to the environment, and genetic drift is the stochastic change in population composition due to random birth and death processes. Migration is the movement of genetic material between spatial locations. The strength of these different evolutionary forces can vary over orders of magnitude, leading to interesting and complex dynamics at the population level [19].

A large body of theoretical work in the field of population genetics has explored each of these evolutionary processes, and the interactions between them [18] - although there still remain many interesting and open theoretical questions, as the complexity of some dynamics makes them challenging to study. Many empirical studies have also tested some theoretical predictions, but limitations in empirical methods have made it challenging to test other predictions. In particular, detecting stochasticity is challenging due to the presence of both biological and technical (non-biological) noise. As a result, there are still many open questions about how the stochastic force of genetic drift acts, and what controls its strength.

In this thesis, we tackle some of these challenges by developing new empirical methods for measuring the strength of genetic drift in a variety of systems, ranging from laboratory experiments to natural populations. This Introduction chapter presents an overview of ge-

netic drift, its consequences for evolution, and current methods and challenges for inferring the strength of genetic drift from data. Additionally, we present a brief background on spatially-structured populations, and in particular range expansions, which are a focus for the work presented in Chapter 2. Finally, we give a brief overview of the work presented in this thesis.

1.1 Genetic drift and its consequences for evolutionary dynamics

Genetic drift is the stochastic change in population composition due to random birth and death processes. The simplest mathematical model of genetic drift is the Wright-Fisher model [48, 14]. The Wright-Fisher model describes a population with discrete non-overlapping generations, a constant population size, and offspring determined by sampling with replacement from the previous generation. While this model makes strong simplifying assumptions, these assumptions are helpful for gaining intuition, and as we see below can also sometimes be used to describe realistic populations using effective parameters. Suppose there are two genotypes, wild type and mutant, in a population with N individuals. The number of mutants, n_{t+1}, of the mutant in generation $t+1$ is given by binomial sampling

$$n_{t+1} = \text{Binom}(N, f_t) \tag{1.1}$$

where $f_t = \frac{n_t}{N}$ is the frequency of the mutant in generation t (Figure 1.1a).

When the population size is large enough such that time and frequency can be treated as continuous variables, t and f, the change in the frequency of the mutant over time can be written as

$$\frac{df}{dt} = \sqrt{\frac{f(1-f)}{N}}\eta(t) \tag{1.2}$$

where $f \equiv f(t)$ is a time-dependent frequency, and $\eta(t)$ is a random variable that is Gaussian-distributed with a mean of 0, variance of 1, and is uncorrelated in time. Equivalently, the dynamics can be described by the Fokker-Planck equation which describes the probability distribution, $p(f,t)$, of frequencies across many instantiations of the population [27, 10]

$$\frac{dp(f,t)}{dt} = \frac{\partial^2}{\partial f^2}\left[\frac{f(1-f)p(f,t)}{2N}\right]. \tag{1.3}$$

In reality, a population will not follow the assumptions of Wright-Fisher dynamics, but it is still convenient (and useful) to compare populations to one another using the Wright-Fisher model. However, because some or all of the assumptions above are broken, the population size is replaced by an effective population size, N_e,

$$\frac{df}{dt} = \sqrt{\frac{f(1-f)}{N_e}}\eta(t). \tag{1.4}$$

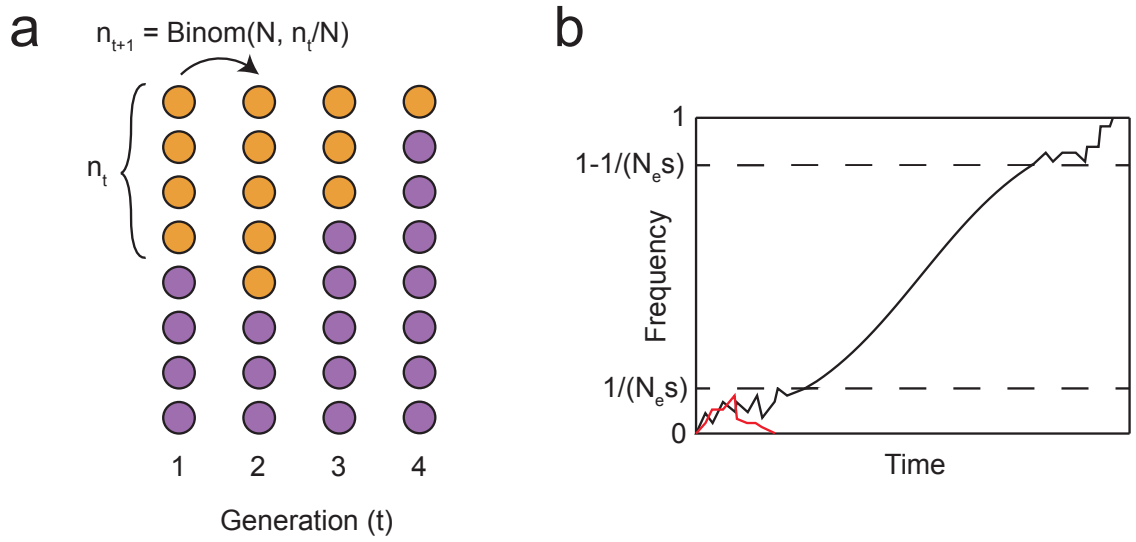

Figure 1.1: (a) The Wright-Fisher model of genetic drift. Offspring are sampled from the previous generation following binomial sampling. (b) The strength of genetic drift dominates over that of selection at low frequencies ($f \lesssim \frac{1}{N_e s}$), causing some weak beneficial mutants to go extinct due to chance (red line), and at high frequencies ($f \gtrsim 1 - \frac{1}{N_e|s|}$), causing some weak deleterious mutants to fix due to chance. Only when a beneficial mutation stochastically fluctuates to above f_{thresh} does the strength of selection dominate (black line).

The effective population size is the population size that would reproduce the observed data in an idealized Wright-Fisher model [7].

When the deviation from Wright-Fisher dynamics is due to additional variation in birth and death processes beyond random sampling of offspring with replacement from the previous generation, the effective population size is given by the true population size divided by the variance in the number of offspring produced by each parent, σ^2,

$$N_e = \frac{N}{\sigma^2}. \tag{1.5}$$

This model is valid as long as the variance in offspring number is finite. Otherwise, a generalized Wright-Fisher model can be used where there is an intermediate step of sampling the distribution of offspring numbers, and then random sampling of this intermediate population to the desired population size in the next generation [36]. Thus, we see that in reality, there are some populations that cannot be described by Wright-Fisher dynamics even with an effective population size, but an effective population size is still conventionally (incorrectly) measured for ease of comparison with other studies [7, 8].

The strength of genetic drift can have a substantial impact on the evolutionary fate of a population. One consequence of genetic drift for evolution is the loss of genetic diversity

over time due to the stochastic extinction of genotypes from the population. The frequencies 0 and 1 are absorbing states; when the frequency of a genotype stochastically fluctuates to 0, it is lost from the population (until it is stochastically created again by mutation or recombination). Populations with a lower effective population size experience faster loss of diversity due to the increased stochasticity from having smaller numbers.

Genetic drift also leads to a weakened efficacy of natural selection. When a beneficial mutation is rare, the strength of genetic drift can dominate over the strength of selection, and the mutation may be lost due to stochastic fluctuations. Only when a mutation fluctuates to a high enough frequency by chance does the strength of selection overpower the strength of genetic drift. Selection can be incorporated into the above equation as a logistic growth term

$$\frac{df}{dt} = sf(1-f) + \sqrt{\frac{f(1-f)}{N_e}}\eta(t) \tag{1.6}$$

where s is the fitness difference of the mutant compared to the wild type (i.e. $s = 0.1$ indicates that the mutant reproduces 10% faster than the wild type). The frequency threshold f_{thresh} above which selection dominates over drift can be heuristically shown from the above equation by equating the the amount of time that it takes for selection or genetic drift to cause a mutant arising at (approximately) frequency 0 to reach frequency f_{thresh} under the simplifying assumption that $f \ll 1$:

$$\delta f_{\text{selection}} = f_{\text{thresh}} = sf_{\text{thresh}}\delta t \tag{1.7}$$

$$\delta f_{\text{drift}} = f_{\text{thresh}} = \sqrt{\frac{f_{\text{thresh}}}{N_e}\delta t} \tag{1.8}$$

$$f_{\text{thresh}} = \frac{1}{N_e s}. \tag{1.9}$$

The probability for a beneficial mutant below f_{thresh} to stochastically fluctuate to f_{thresh}, or establish, is thus given by

$$p_{\text{est}} = N_e s f_0 \tag{1.10}$$

where f_0 is the initial frequency of the mutant. Thus, weak beneficial mutations can be lost stochastically due to chance (Figure 1.1b). Additionally, weak deleterious mutations ($s < 0$) have a non-zero chance of fixing if they start above a frequency of $f_0 = 1 - \frac{1}{N_e|s|}$ and they stochastically fluctuate to fixation due to genetic drift. The fixation probability for these deleterious mutants is

$$p_{\text{fix}} = N_e|s|(1-f_0). \tag{1.11}$$

A variety of factors influence the strength of genetic drift including population size, bottleneck size, the offspring number distribution, and population structure (for instance see review [7]). Lower population sizes lead to lower effective population sizes, larger frequency fluctuations, and a higher strength of genetic drift. A population may also have a bottleneck in the population size, which is a decrease in population size, for instance when a pathogen

is transmitted between hosts, which leads to a decreased effective population size. As mentioned above, variability in the number of offspring produced also leads to a low effective population size and increased genetic drift. Population structure may arise due to spatial segregation, or segregation into other demographic groups such as by age, and affect the effective population size.

Outside of the reproduction and death of individuals in a population, genetic drift can also arise in pathogen transmission. In pathogen transmission, births are caused by transmission of the pathogen to a different host and deaths are caused by recovery or death of an infected host. Thus, the theoretical framework described above can also be used to study between-host evolutionary dynamics of pathogens.

1.2 Observables and methods for measuring genetic drift

Due to the importance of genetic drift for evolutionary dynamics, a variety of methods have been developed to infer the strength of genetic drift from data, and we will summarize the broad classes of these methods here. The particular method that is chosen depends on the data that is available and the goals of the study.

One observable that can be used to infer the strength of genetic drift is the rate of diversity loss. A population with a higher level of genetic drift has a faster loss of diversity, given that the initial diversity and the rate of other evolutionary processes is the same between the two populations. For instance, methods have been developed to infer bottleneck sizes of between-host pathogen transmission using diversity loss as a metric [40], and these methods have been applied to study influenza [33] and SARS-CoV-2 [31, 32], and other pathogens.

More generally, genetic drift sets the rate at which the variance of the change in frequency increases over time. The variance in the frequency change, δf, across a single generation is given by

$$\mathrm{var}(\delta f) = \frac{f(1-f)}{N_e}. \tag{1.12}$$

Thus by measuring the variance of the distribution of frequency changes, the strength of genetic drift can be inferred. This approach was used in one of the first experimental measurements of genetic drift, which was performed on populations of fruit flies with different eye colors [6]. By tracking the frequency of flies with each eye color over multiple generations for a large number of replicate populations, the effective population size of the fly population could be quantified (Figure 1.2a). Variations of methods that track frequency changes over time to infer effective population size have been developed since then to account for additional complexities such as empirical sampling and the joint inference of fitness and effective population size [4, 12, 13, 5, 44].

Another class of methods that is used to infer the strength of genetic drift uses the phylogenetic tree. First, genomic sequences from the population are used to construct a

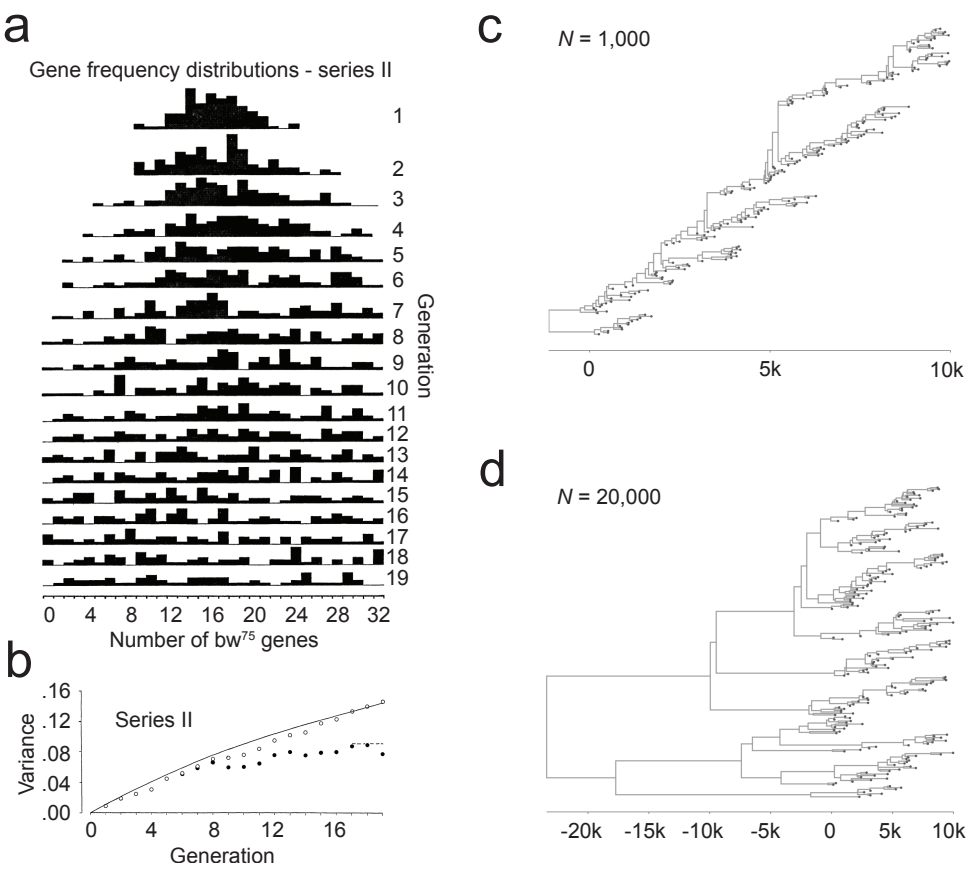

Figure 1.2: (a) The distribution of the number number of bw^{75} genes in a fly population, for 60 populations, as a function of the generation number. The variance of the distribution increases over time due to genetic drift. Adapted and reproduced with permission from Ref. [6]. (b) The variance of the change in frequency of bw^{75} genes over time across the same 60 populations as in (a). Open circles show the observed variance including previously fixed populations (frequency of 0 or 1). Closed circles show the observed variance excluding previously fixed populations. The black line shows the theoretical expectation, $\text{var}(\Delta f) = f_0(1-f_0)[1-(1-\frac{1}{2N_e})^n]$, where n is the number of generations [49]. This form results from integrating Equation 1.12 across multiple generations, and replacing N_e by $2N_e$ because flies are a diploid population. $2N_e$ for the black line is taken to be 23. Adapted and reproduced with permission from Ref. [6]. (c-d) The phylogenetic tree for a simulated population with $N = 1000$ and $N = 20,000$ showing that a population with a lower population size exhibits faster coalescence times between any two extant lineages, going backwards in time. Adapted and reproduced from Ref. [2] which is licensed under a Create Commons Attribution 2.0 Generic License (https://creativecommons.org/licenses/by/2.0/).

tree based on genetic similarity (a phylogenetic tree). Then a model of the rate at which branches in the tree merge going backwards in time (the coalescence rate), can be fit to the observed tree to infer the effective population size. The most common model that is used for the coalescence rate is the Kingman coalescent, which describes populations that can be described by Wright-Fisher dynamics [28]. The rate λ at which lineages coalesce backwards in time in the Kingman coalescent is given by

$$\lambda(i) = \frac{i(i-1)}{2N_e\tau} \tag{1.13}$$

where i is the number of lineages present at a given time, N_e is the effective population size, and τ is the generation time. In this model, populations with higher levels of genetic drift (lower effective population size) tend to have a higher coalescence rate (Figure 1.2b). Additionally the rate of coalescence decreases going backwards in time due to the decrease in the number of lineages (lower i). A variety of methods have been developed to fit the coalescence rate from observed trees with different priors on how N_e changes over time [37, 25, 46]. Methods have also been developed using models that differ from the Kingman coalescent [41]. These phylogenetics methods have been used extensively to study microbial evolution as well as the evolution of macroscopic organisms like humans.

Branching process models can also be fit to the observed number of individuals over time to infer the variance in offspring number and growth rate [9]. These models assume a particular distribution for the offspring number distribution (often a negative binomial distribution) and estimate the most likely combination of parameters of the offspring number distribution along with growth rate. External information about the growth rate can be used to then determine the parameters of the offspring number distribution.

Finally, the offspring number distribution and population size can be directly measured and used to calculate the effective population size. For instance, in epidemics, contact tracing can be used to measure the distribution of secondary cases from an infected individual [29, 42, 1, 3], and surveys can be used to determine the disease prevalence [15] (approximately equivalent to the population size N in the models described above). In the absence of direct disease contact tracing data, information about the general rate of contacts between different groups can be used in combination with modeling to infer the distribution of offspring numbers [38]. In laboratory evolution experiments, the distribution of offspring numbers can be measured directly using microscopy [30] or by using sequencing to measure the distribution of genotype abundances [47].

1.3 Microbial range expansions

Range expansions occur when populations expand into an area of space that they previously did not inhabit. Microbial range expansions commonly occur in biofilms, which are populations of microbes (bacteria or yeast) that concentrate at interfaces. Biofilms are highly prevalent in nature, such as in human-associated microbial infections, natural hot springs,

and in biofouling. In biofilms, cells commonly create an extracellular matrix, which is a sticky substance composed of polysaccharides, proteins, and nucleic acids that helps cells stick to each other [21].

Evolutionary dynamics in microbial range expansions have been studied extensively, primarily using microbial colonies as a model system. When a microbial colony grows large enough in size, nutrient limitation causes faster growth at the leading edge of the colony compared to the center, leading to interesting evolutionary dynamics that differ from those in a well-mixed environment [26]. For instance, range expansions can promote adaptation by increasing the number of generations to reach a final population size as a result of nutrient limitation, thus increasing the amount of time over which selection will act. Additionally, genetic drift will lead to the stochastic extinction of genotypes from the leading edge of the colony, which causes the formation of sectors, or regions that are composed of a single genotype (Figure 1.3a) [24, 23]. If a mutation happens to occur at the expanding frontier of the colony and is lucky to escape extinction due to genetic drift, it will become a sector, which is a process called "gene surfing". These sectors are "jackpots", or clones (descendants of a particular ancestor) that happened to reach a large size [17]. Interestingly, gene surfing leads to offspring numbers with a heavy-tail where the Wright-Fisher model cannot be applied [22].

Additionally, whereas in well-mixed culture, in the absence of other evolutionary forces, genetic drift will lead to the eventual take-over of a single genotype in the population, in range expansions, the maintenance of diversity is possible in a circular range expansion because the rate at which the population size grows balances the rate at which genetic drift causes the extinction of genotypes [23]. An additional consequence of genetic drift in range expansions that is not found in well-mixed culture is the ability to promote cooperation. The formation of sectors leads to the spatial proximity of individuals from the same genotype. As a result, cooperation is beneficial and cheaters can be excluded [35].

A variety of observables can be used to infer the strength of genetic drift in range expansions. Because the Wright-Fisher model assumptions can be broken in range expansions, determining an effective population size is not very meaningful. Instead, the rate of change of the number of sectors N over time in a colony with standing variation can be fit to the theory expectation [23]. By assuming that the radius of the colony grows linearly after a transient exponential growth period and before complete nutrient depletion, the colony radius r can be used as a proxy for time. The mean number of sectors as a function of colony radius is given by

$$N(r|r_0) = \sqrt{\frac{\pi}{2D_X(\frac{1}{r_0} - \frac{1}{r})}} \tag{1.14}$$

where r_0 is the initial radius of the colony and D_X is the diffusion constant of the sector boundary that divides two genotypes (Figure 1.3b). From this equation, we see that the diffusion constant of the sector boundary can also be measured to determine the strength of genetic drift, where a higher diffusion constant corresponds to a faster rate of diversity loss and thus a higher strength of genetic drift. It has also been shown that the diffusion

constant of single cell trajectories very closely mirrors that of sector boundaries [20], and thus can also be used as an observable for measuring the strength of genetic drift. In reality, the movement of single cells or sector boundaries often exhibits superdiffusion rather than normal diffusion, and Equation 1.14 can be modified accordingly, as is done in Ref. [23].

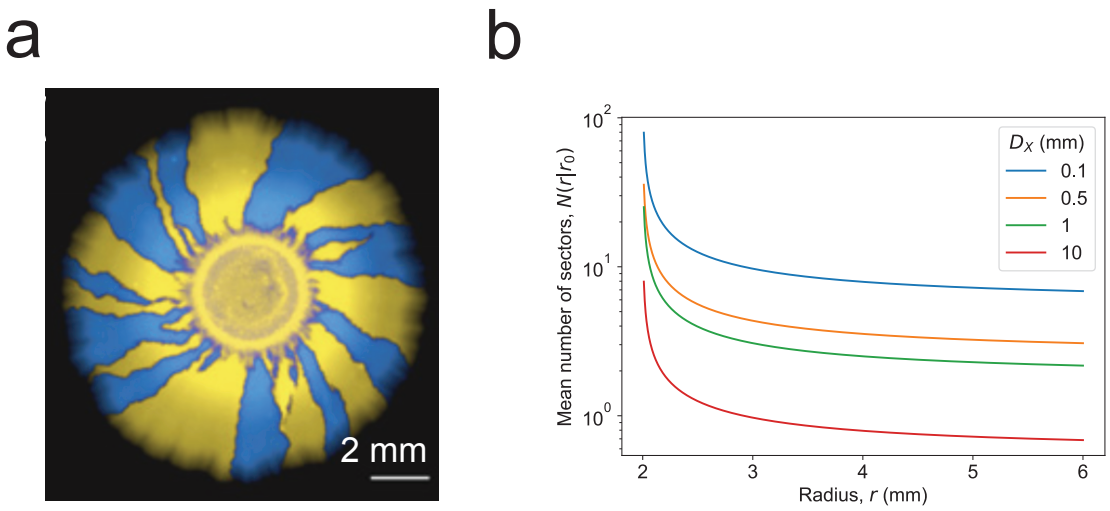

Figure 1.3: (a) *E. coli* DH5α colony grown from an equal mixture of neutral strains expressing yellow fluorescent protein and cyan fluorescent protein on a plasmid, grown at 37°C. Adapted and reproduced with permission from Ref. [20]. (b) The mean number of sectors in an expanding microbial colony as a function of the radius of the colony, given by Equation 1.14, where $r_0 = 2$ mm. D_X is the diffusion constant of the sector boundaries separating the two genotypes.

1.4 Limitations in our current understanding of genetic drift

Despite decades of theoretical and experimental research on genetic drift, there are still many limitations in our understanding of how this evolutionary process works. The strength of genetic drift is often considered to be a static property of a population on short timescales rather than an evolvable trait. However, we have some evidence that mutations can change the strength of genetic drift, for instance for particular mutations of *E. coli* in microbial colonies [20]. Other evolutionary processes such as natural selection and mutation can be evolvable traits of a population (i.e. mutations that change fitness or mutations in the DNA repair pathways that change the mutation rate [39, 43]). Thus, it is natural to hypothesize

that genetic drift can also be an evolvable trait of a population. This hypothesis has not yet been explored.

Additionally, we have limited empirical measurements of how microscopic properties at the individual level lead to genetic drift at the population level. While this question has been explored to some degree in microbial range expansions both experimentally and computationally [34, 35, 11, 24, 26], experiments have been limited by throughput and theoretical models do not necessarily completely capture the growth dynamics. By studying the connection between individual-level traits and population-level genetic drift, we can begin to gain a better understanding of what causes changes to genetic drift in microbial populations.

These limitations of our understanding are due in part to methodological challenges. One of the main challenges to inferring the strength of genetic drift is that current methods do not scale well to high-throughput applications. Current methods require experimentally tagging different strains with fluorescent or genetic markers; this can be time-consuming and can realistically only be done for a handful of strains [20]. Another main challenge is how to incorporate measurement or technical noise, which is noise that is due to the measurement process rather than biological processes [16]. As both measurement noise and genetic drift are stochastic, they may create signals in the data that are confounding. Without correcting for measurement noise, the inferred strength of genetic drift may be biased.

In this thesis, we address these limitations by developing improved experimental and computational methods to infer genetic drift from data. These methods allow us to determine how the strength of genetic drift is affected by mutations, and if so, what consequence it has for evolution. Additionally, our improved computational methods allow inference of genetic drift in pathogen transmission using large pathogen genomics datasets while accounting for noise in the measurement process. Finally, we empirically measure phenotypic traits at both the individual and population levels, and then we model their effects on genetic drift at the population level.

1.5 Outline of thesis

In Chapter 2, we test the hypothesis that mutations can change the strength of genetic drift, the first requirement needed for a trait to be evolvable. We focus on microbial colonies, which are a model system for range expansions. To test this hypothesis, we develop a new experimental method to measure the strength of genetic drift in high throughput using fluorescence microscopy. We find that mutations significantly affect the strength of genetic drift by causing changes in the self-organized spatial structure of the colony. These changes to genetic drift substantially affect the probability that new beneficial mutations escape stochastic extinction, providing evidence that genetic drift may be an evolvable property of a population.

In Chapter 3, we investigate the strength of genetic drift in SARS-CoV-2 evolution at the host population level. We develop a statistical inference framework for inferring the strength

of genetic drift simultaneously with measurement noise from lineage frequency time series data. Applying this method to genomic data from England, we find that the strength of genetic drift is consistently, throughout time, higher than expected given the number of people infected with COVID-19 at the community level, even after correcting for measurement noise and epidemiological dynamics. We also find evidence for spatial structure in SARS-CoV-2 transmission at the regional level. The levels of genetic drift that we observe are higher than the estimated levels of superspreading found by modeling studies that incorporate data of contact statistics in England. We discuss how even in the absence of superspreading, high levels of genetic drift can be generated via jackpot events in a deme model.

1.6 Bibliography

[1] D. C. Adam, P. Wu, J. Y. Wong, E. H. Lau, T. K. Tsang, S. Cauchemez, G. M. Leung, and B. J. Cowling. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine*, 26(11):1714–1719, 2020.

[2] T. Bedford, S. Cobey, and M. Pascual. Strength and tempo of selection revealed in viral gene genealogies. *BMC Evolutionary Biology*, 11(1):1–16, 2011.

[3] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*, 20(8):911–919, 2020.

[4] J. P. Bollback, T. L. York, and R. Nielsen. Estimation of $2N_e s$ from temporal allele frequency data. *Genetics*, 179(1):497–502, 2008.

[5] V. Buffalo and G. Coop. The linked selection signature of rapid adaptation in temporal genomic data. *Genetics*, 213(3):1007–1045, 2019.

[6] P. Buri. Gene frequency in small populations of mutant Drosophila. *Evolution*, pages 367–402, 1956.

[7] B. Charlesworth. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3):195–205, 2009.

[8] R. Der, C. L. Epstein, and J. B. Plotkin. Generalized population models and the nature of genetic drift. *Theoretical Population Biology*, 80(2):80–99, 2011.

[9] A. Endo et al. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, 5, 2020.

[10] W. J. Ewens. *Mathematical Population Genetics: Theoretical Introduction*, volume 1. Springer, 2004.

[11] F. D. Farrell, M. Gralka, O. Hallatschek, and B. Waclaw. Mechanical interactions in bacterial colonies and the surfing probability of beneficial mutations. *Journal of The Royal Society Interface*, 14(131):20170073, 2017.

[12] A. F. Feder, S. Kryazhimskiy, and J. B. Plotkin. Identifying signatures of selection in genetic time series. *Genetics*, 196(2):509–522, 2014.

[13] A. Ferrer-Admetlla, C. Leuenberger, J. D. Jensen, and D. Wegmann. An approximate Markov model for the Wright–Fisher diffusion and its application to time series data. *Genetics*, 203(2):831–846, 2016.

[14] R. A. Fisher. *The Genetical Theory of Natural Selection*. Dover Publications, 1958.

[15] U. O. for National Statistics. Coronavirus (COVID-19) Infection Survey: England. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata. Accessed: 2021-12-10.

[16] S. D. Frost, O. G. Pybus, J. R. Gog, C. Viboud, S. Bonhoeffer, and T. Bedford. Eight challenges in phylodynamic inference. *Epidemics*, 10:88–92, 2015.

[17] D. Fusco, M. Gralka, J. Kayser, A. Anderson, and O. Hallatschek. Excess of mutational jackpot events in expanding populations revealed by spatial Luria–Delbrück experiments. *Nature Communications*, 7(1):1–9, 2016.

[18] J. H. Gillespie. *Population Genetics: A Concise Guide*. JHU press, 2004.

[19] B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678):45–50, 2017.

[20] M. Gralka, F. Stiewe, F. Farrell, W. Möbius, B. Waclaw, and O. Hallatschek. Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters*, 19(8): 889–898, 2016.

[21] L. Hall-Stoodley, J. W. Costerton, and P. Stoodley. Bacterial biofilms: from the natural environment to infectious diseases. *Nature Reviews Microbiology*, 2(2):95–108, 2004.

[22] O. Hallatschek and D. R. Nelson. Gene surfing in expanding populations. *Theoretical Population Biology*, 73(1):158–170, 2008.

[23] O. Hallatschek and D. R. Nelson. Life at the front of an expanding population. *Evolution*, 64(1):193–206, 2010.

[24] O. Hallatschek, P. Hersen, S. Ramanathan, and D. R. Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930, 2007.

[25] S. Y. Ho and B. Shapiro. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, 11(3):423–434, 2011.

[26] J. Kayser, C. F. Schreck, Q. Yu, M. Gralka, and O. Hallatschek. Emergence of evolutionary driving forces in pattern-forming microbial populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1747):20170106, 2018.

[27] M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713, 1962.

[28] J. F. Kingman. Origins of the coalescent: 1974-1982. *Genetics*, 156(4):1461–1463, 2000.

[29] R. Laxminarayan, B. Wahl, S. R. Dudala, K. Gopal, C. Mohan B, S. Neelima, K. Jawahar Reddy, J. Radhakrishnan, and J. A. Lewnard. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science*, 370(6517):691–697, 2020.

[30] I. Levin-Reisman, O. Gefen, O. Fridman, I. Ronin, D. Shwa, H. Sheftel, and N. Q. Balaban. Automated imaging with scanlag reveals previously undetectable bacterial growth phenotypes. *Nature Methods*, 7(9):737–739, 2010.

[31] K. A. Lythgoe, M. Hall, L. Ferretti, M. de Cesare, G. MacIntyre-Cockett, A. Trebes, M. Andersson, N. Otecko, E. L. Wise, N. Moore, et al. SARS-CoV-2 within-host diversity and transmission. *Science*, 372(6539):eabg0821, 2021.

[32] M. A. Martin and K. Koelle. Comment on "Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2". *Science Translational Medicine*, 13(617):eabh1803, 2021.

[33] J. T. McCrone, R. J. Woods, E. T. Martin, R. E. Malosh, A. S. Monto, and A. S. Lauring. Stochastic processes constrain the within and between host evolution of influenza virus. *eLife*, 7:e35962, 2018.

[34] S. Mitri, E. Clarke, and K. R. Foster. Resource limitation drives spatial organization in microbial groups. *The ISME Journal*, 10(6):1471–1482, 2016.

[35] C. D. Nadell, K. R. Foster, and J. B. Xavier. Emergence of spatial structure in cell groups and the evolution of cooperation. *PLoS Computational Biology*, 6(3):e1000716, 2010.

[36] T. Okada and O. Hallatschek. Dynamic sampling bias and overdispersion induced by skewed offspring distributions. *Genetics*, 219(4):iyab135, 2021.

[37] O. G. Pybus, A. Rambaut, and P. H. Harvey. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, 155(3):1429–1437, 2000.

[38] B. J. Quilty, L. A. Chapman, K. L. Wong, A. Gimma, S. Pickering, S. JD, R. P. G. Neil, C. I. Jarvis, and A. J. Kucharski. Reconstructing the secondary case distribution of SARS-CoV-2 from heterogeneity in viral load trajectories and social contacts. *Report for SPI-M-O and SAGE*, 2021.

[39] P. D. Sniegowski, P. J. Gerrish, and R. E. Lenski. Evolution of high mutation rates in experimental populations of E. coli. *Nature*, 387(6634):703–705, 1997.

[40] A. Sobel Leonard, D. B. Weissman, B. Greenbaum, E. Ghedin, and K. Koelle. Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. *Journal of Virology*, 91(14):e00171–17, 2017.

[41] T. Stadler. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1):58–66, 2009.

[42] K. Sun, W. Wang, L. Gao, Y. Wang, K. Luo, L. Ren, Z. Zhan, X. Chen, S. Zhao, Y. Huang, et al. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science*, 371(6526):eabe2424, 2021.

[43] F. Taddei, M. Radman, J. Maynard-Smith, B. Toupance, P.-H. Gouyon, and B. Godelle. Role of mutator alleles in adaptive evolution. *Nature*, 387(6634):700–702, 1997.

[44] P. Tataru, M. Simonsen, T. Bataillon, and A. Hobolth. Statistical inference in the Wright–Fisher model using allele frequency data. *Systematic Biology*, 66(1):e30–e46, 2017.

[45] L. A. Urry, M. L. Cain, S. A. Wasserman, P. V. Minorsky, and R. Orr. *Campbell Biology, 12th Edition*. Pearson, 2021.

[46] E. M. Volz and X. Didelot. Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. *Systematic Biology*, 67(4):719–728, 2018.

[47] E. S. Wright and K. H. Vetsigian. Stochastic exits from dormancy give rise to heavy-tailed distributions of descendants in bacterial populations. *Molecular Ecology*, 28(17): 3915–3928, 2019.

[48] S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97, 1931.

[49] S. Wright. Statistical genetics and evolution. *Bulletin of the American Mathematical Society*, 48(4):223–246, 1942.

Chapter 2

Mutability of demographic noise in microbial range expansions

Genetic drift is often considered a static property of a population on short timescales rather than an evolvable trait. However, we have evidence that particular single mutations can substantially change the strength of genetic drift in microbial range expansion [26]. Here we test the hypothesis of whether random single gene deletion mutations (a common type of single step mutation in bacteria) can substantially impact the strength of genetic drift in microbial range expansions. Furthermore, we ask what phenotypic traits determine the strength of genetic drift, and whether the changes to genetic drift due to single mutations substantially affect downstream evolution.

I am grateful to Matti Gralka for performing preliminary experiments of the strength of genetic drift in selected strains from the Keio collection, Marie-Cécilia Duvernoy for advising on general project direction and guidance with imaging and image analysis, Megan Sousa for performing the experiments and analysis of single cell shape in different growth environments, Arbel Harpak for advising on statistical analyses, Oskar Hallatschek for general advising, and all authors for helpful feedback on the work and writing presented in this chapter. The remainder of this chapter was published in its entirety as

> QinQin Yu, Matti Gralka, Marie-Cécilia Duvernoy, Megan Sousa, Arbel Harpak, Oskar Hallatschek. Mutability of demographic noise in microbial range expansions. The ISME Journal **15**, 2643–2654 (2021). https://doi.org/10.1038/s41396-021-00951-9

2.1 Abstract

Demographic noise, the change in the composition of a population due to random birth and death events, is an important driving force in evolution because it reduces the efficacy of natural selection. Demographic noise is typically thought to be set by the population size and the environment, but recent experiments with microbial range expansions have revealed substantial strain-level differences in demographic noise under the same growth conditions. Many genetic and phenotypic differences exist between strains; to what extent do single mutations change the strength of demographic noise? To investigate this question, we developed a high-throughput method for measuring demographic noise in colonies without the need for genetic manipulation. By applying this method to 191 randomly-selected single gene deletion strains from the *E. coli* Keio collection, we find that a typical single gene deletion mutation decreases demographic noise by 8% (maximal decrease: 81%). We find that the strength of demographic noise is an emergent trait at the population level that can be predicted by colony-level traits but not cell-level traits. The observed differences in demographic noise from single gene deletions can increase the establishment probability of beneficial mutations by almost an order of magnitude (compared to in the wild type). Our results show that single mutations can substantially alter adaptation through their effects on demographic noise and suggest that demographic noise can be an evolvable phenotype of a population.

2.2 Introduction

Demographic noise, also referred to as "genetic drift", "neutral drift", or "drift", is the change in the composition of a population due to random births and deaths. Theoretical population genetics predicts that demographic noise competes with natural selection by lowering the establishment probability of beneficial mutations [35] and causing the accumulation of deleterious mutations [36, 51], leading to consequences such as the existence of a drift barrier [34] (a minimum absolute value fitness above which selection can act) and Muller's ratchet [47]. Additionally, demographic noise reduces neutral genetic diversity [30], can limit mutation rates [43], and can also promote cooperation in spatially-structured environments [49]. Experimental studies have validated many of these predictions [11, 3, 10], and demographic noise has been shown to play an important role in the evolutionary dynamics of a variety of systems including organelles [44], intestinal crypt stem cells [42], biofilms [59], the transmission of viruses [18, 45, 55] and human mitochondrial DNA [65], well-mixed culture [29], and potentially some types of cancer tumors [63].

Intuitively, the randomness of individual birth and death events should matter only relative to the population's size (which can be influenced by the environment), which is conventionally thought to set the strength of demographic noise [13, 21, 25, 19, 31, 26, 46, 8, 7, 4]. However, recent work in microbial colonies has shown that different strains from the same species can exhibit different strengths of demographic noise under the same

growth conditions [26, 28], and that the observed differences in demographic noise can have a substantial impact on the establishment probability of beneficial mutations [26, 14, 20]. However, it is unknown how much single mutations can affect the strength of demographic noise and whether those changes would be sufficient to alter the efficacy of natural selection. In this work, we focus on loss of function mutations using single gene deletion mutant strains, as loss of function mutations are a common type of single step mutation in microbes.

Measuring the strength of demographic noise for a large number of strains requires a method for high-throughput tracking of cellular lineages in growing colonies. Previous methods for measuring demographic noise in microbial colonies required genetic transformations [28] or time-intensive microscopy and image analysis [26], which are impractical for testing a large number of strains. Here, we develop a label-free method to sparsely track cell lineages (i.e. at low density) in growing colonies and use it to measure the distribution of demographic noise effects in *E. coli* single gene deletion strains. We show that most gene deletions decrease the strength of demographic noise, which in turn can dramatically increase the establishment probability of beneficial mutations. Our high-throughput approach also allows us to show that population-level emergent properties such as colony shape and size, but not single-cell properties such as cell shape, can predict the strength of demographic noise.

2.3 Methods

Strains and growth conditions

Single gene deletion strains were taken from the Keio collection [5] (Supplementary table 1), which consists of all non-essential single gene deletions in *E. coli* K-12 strain BW25113. *mreB* and *mrdA* point mutant strains were from Ref [57] (Supplementary table 2). Plasmids pQY10 and pQY11 were created by Gibson assembly of Venus YFP A206K (for pQY10) or Venus CFP A206K (for pQY11) [28], and *specR* from pKDsgRNA-ack (gift from Kristala Prather, Addgene plasmid # 62654, http://n2t.net/addgene:62654 ; RRID:Addgene_62654) [54]. Plasmids pQY12 and pQY13 were created similarly but additionally with *cmR* from pACYC184.

All *E. coli* experiments were performed in LB (Merck 110285, Kenilworth, New Jersey) with the appropriate antibiotics and experiments with *S. cerevisae* were performed in YPD [2]. All agar plates were prepared in OmniTrays (Nunc 242811, Roskilde, Denmark, 12.8cm×8.6cm) or 12cm×12cm square petri dishes (Greiner 688102, Kremsmuenster, Austria) filled with 70 mL media solidified with 2% Bacto Agar (BD 214010, Franklin Lakes, New Jersey). After solidifying, the plates were dried upside-down in the dark for 2 days and stored wrapped at 4°C in the dark for 7-20 days before using.

Tracking lineages with fluorescent tracer beads

In order to track lineages, we spread fluorescent tracer beads with a similar size to the cells on the surface of an agar plate, allowed them to dry, then inoculated and grew a colony on top of the agar plate and imaged the tracer beads to track lineages. In this way, we are able to track lineages without genetic labels at low density (i.e. sparsely) in the colony so that we can distinguish individual lineages without needing high-resolution microscopy. We find that the bead trajectories track cell lineages over the course of one hour both at the colony front and behind the front (Figures 2.1c, 2.S1c-d, and 2.S2). We chose to spread fluorescent tracer beads on the surface of the agar so that they could continue to be incorporated into the colony as it grew, which would allow us to track lineages even as existing beads and lineages get lost from the front. Even though many cells will be piled up on top of other cells rather than in contact with the agar, we don't expect this to affect the ability of the beads to measure demographic noise, since lineages at the front are the most likely to contribute offspring to future generations [26].

Fluorescent tracer beads

For experiments with *E. coli*, 1 μm red fluorescent polystyrene beads from Magsphere (PSF-001UM, Pasadena, CA, USA) were diluted to 3 μg/mL in molecular grade water and 920 μL was spread on the surface of the prepared OmniTray agar plates with sterile glass beads. Excess bead solution was poured out, and the plates were dried under the flow of a class II biosafety cabinet (Nuaire, NU-425-300ES, Plymouth, MN, USA) for 45 minutes. The bead density was chosen to achieve $\sim$250 beads in a 56x field of view. For experiments with *S. cerervisiae*, 2μm dragon green fluorescent polystyrene beads from Bang's labs (FSDG005, Fishers, IN, USA) were used at a similar surface density.

Measurement of the distribution of demographic noise

We randomly selected 352 single gene deletion strains from the Keio collection. For each experiment, cells were thawed from glycerol stock (see Supplementary methods), mixed, and 5 μL was transferred into a 96-well flat bottom plate with 100μL LB and the appropriate antibiotics. Plates were covered with Breathe-Easy sealing membrane (Diversified Biotech BEM-1, Doylestown, PA, USA) and grown for 12 hours at 37°C without shaking. A floating pin replicator (V&P Scientific, FP12, 2.36 mm pin diameter, San Diego, CA, USA) was used to inoculate a 2-3mm droplet from each well of the liquid culture onto a prepared OmniTray covered with fluorescent tracer beads. Droplets were dried and the plates were incubated upside down at 37°C for 12 hours before timelapse imaging.

To account for systematic differences between plates, we also put 8 wild type BW25113 wells in each 96-well plate in different positions on each plate. The mean squared displacement (MSD, see below) of each gene deletion colony was normalized to the weighted average MSD of the wild type BW25113 colonies on that plate, $\langle \text{MSD} \rangle_{WT}$, and this "relative MSD"

is reported. We performed three biological replicates for each strain (grown from the same glycerol stock, Figure 2.S3), and their measurements were averaged together weighted by the inverse of the square of their individual error in relative MSD. The reported error for the strain is the standard error of the mean. During the experiment, several experimental challenges impede our ability to measure demographic noise, including the appearance of beneficial sectors (identified as diverging bead trajectories that correspond to bulges at the colony front) either due to de novo beneficial mutations or standing variation from glycerol stock (see Supplementary information "Beneficial sectors in monoculture colonies", Figures 2.S4 and 2.S5), slow growth rate leading to bead tracks that were too short for analysis, no cells transferred during inoculation with our pinning tool, inaccurate particle tracking due to beads being too close together, or out of focus images. In order to keep only the highest quality data points, we focused on the 191 strains that had at least 2 replicates free of such issues.

Timelapse imaging of fluorescent beads

Plates were transferred to an ibidi stagetop incubator (Catalog number 10918, Gräfelfing, Germany) set to 37°C for imaging. Evaporation was minimized by putting wet Kim wipes in the chamber and sealing the chamber with tape. The fluorescent tracer beads at the front of the colony were imaged with a Zeiss Axio Zoom.V16 (Oberkochen, Germany) at 56x magnification. A custom macro program written using the Open Application Development for Zen software was used to find the initial focal position for each colony and adjust for deterministic focus drift over time due to slight evaporation. Timelapse imaging was performed at an interval of 10 minutes for 12 hours, during which time the colony grew about halfway across the field of view. Two z slices were taken for each colony and postprocessed to find the most in-focus image to adjust for additional focus drift. Subpixel-resolution particle tracking of the bead trajectory was achieved using a combination of particle image velocimetry and single particle tracking [12] and is described in detail in the Supplementary methods.

Measurement of bead trajectory mean squared displacement

The measurement of mean squared displacement (MSD) is adapted from [28] and is illustrated in Figure 2.1a and 2.S1a. Points in a trajectory that fall within a window of length L are fit to a line of best fit. The MSD is given by

$$MSD(L) = \left\langle \left\langle \int_{l}^{l+L} (\Delta w(L'))^2 dL' \right\rangle_{windows} \right\rangle_{trajectories} \tag{2.1}$$

where $\Delta w(L')$ is the displacement of the bead trajectory from the line of best fit at each point, $\langle\rangle_{windows}$ is an average over all possible definitions of a window with length L along the trajectory (window definitions are overlapping), and $\langle\rangle_{trajectories}$ is a weighted average over all trajectories in a field of view, where the weight is the inverse squared standard

error of the mean for each trajectory's MSD(L) (Figure 2.S1a). We use 200 linearly spaced window sizes from $L = 6$ to 1152 μm. Window sizes that fit in fewer than 5 trajectories are dropped due to the noisiness in calculating the averaged MSD(L). The combined MSD(L) for all trajectories reflects that of bead trajectories at the colony front, which will have the largest contribution to the strength of demographic noise [26] (Figure 2.S6). Because we expect the trajectories to follow an anomalous random walk [28], the combined MSD(L) for all trajectories across the field of view is fit using weighted least squares to a power law, where the weight is the inverse square of the propagated standard error of the mean. Colonies with data in fewer than 5 window sizes are dropped due to the noisiness in fitting to a power law. The fit is extrapolated to $L = 50$ μm to give a single summary statistic for each colony, and this quantity is reported as MSD(L = 50 μm) (see Supplementary information "Determining the mean squared displacement window size", Figures 2.S7 and 2.S8), and the error is calculated as half the difference in MSD(L = 50μm) from using the upper and lower bounded coefficients to the fit. For Figure 2.2, only MSD values where the error is less than half of the value are kept.

Measurement of phenotypic traits

For the phenotypic trait measurements, in addition to the 191 single gene deletions, we also measured 41 additional strains of *E. coli* which included 4 strain backgrounds, 1 *mreB* knockout in the MC1000 background, 2 adhesin mutants, and 34 single gene knockouts from the Keio collection that we predicted may have large changes to demographic noise because of an altered biofilm forming ability in liquid culture [50] or altered cell shape from the wild type (using the classification on the Keio website, `https://shigen.nig.ac.jp/ecoli/strain/resource/keioCollection/list`). We normalized all phenotypic trait values to the average value measured from the wild type colonies on the same plate. The reported values for each strain are averages across 2-3 replicate colonies on different plates and the errors are the standard error of the mean. See the Supplementary methods for more details of the specific phenotypic trait measurements.

Measurement of neutral fraction of diversity preserved

Neutral fluorescent pairs were created by transforming background strains with plasmids pQY10 (YFP, *specR*) or pQY11 (CFP, *specR*). Cells were streaked from glycerol stock and a single colony of each strain was inoculated into a 96 well plate with 600 μL LB and 120 μg/mL spectinomycin for plasmid retention. Plates were covered with Breathe-Easy sealing membrane and grown for 12 hours at 37°C without shaking. 50 μL of culture from each strain in a neutral pair were mixed and a floating pin replicator was used to inoculate a 2-3mm droplet from the liquid culture onto a prepared OmniTray covered with fluorescent tracer beads. Droplets were dried and the plates were incubated at 37°C.

Colonies were imaged after 24 hours with fluorescence microscopy using a Zeiss Axio Zoom.V16 and the number of sectors of each color was manually counted. The fraction of

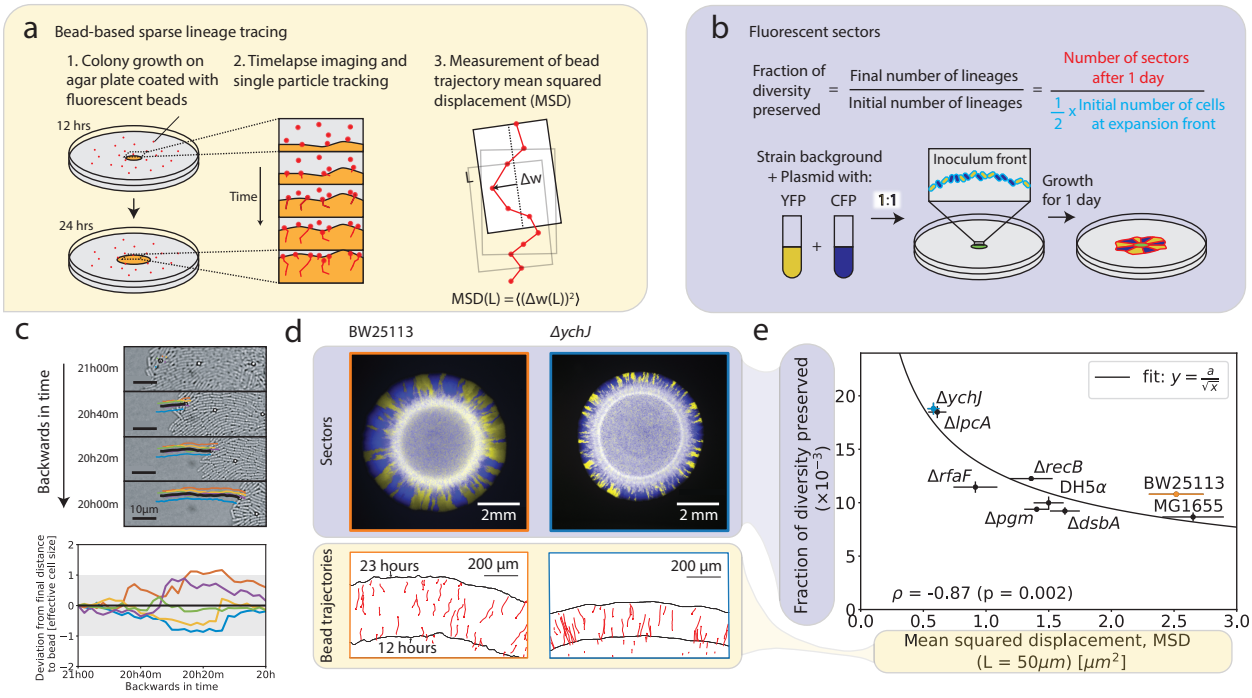

Figure 2.1: Label-free method of measuring demographic noise in microbial colonies (a) Schematic of bead-based sparse lineage tracing method for measuring demographic noise. (b) Schematic of existing method for measuring fraction of diversity preserved [26]. (c) (Top) The trajectory of a single bead (black) and the lineages of the cells neighboring it in the final-timepoint (colors) traced backwards in time in the Keio collection wild type strain. (Bottom) The deviation of the distance between the cell lineages and the bead from the final distance, backwards in time. Colors are the same as in the time series images. The gray shaded region shows a single cell width away or towards the bead. All cells that neighbor the bead in the final timepoint, except for one (orange), are neighbors of the bead in the first timepoint and stay within a single cell width of the final distance to the bead. (d) Example neutral mixtures of YFP and CFP tagged strains grown for 1 day and bead trajectories for strains highlighted in (e). (e) Comparison of MSD at window size L = 50 μm to the fraction of diversity preserved for 3 *E. coli* strain backgrounds and 6 single gene deletions on the Keio collection wild type background (BW25113). Error bars in MSD represent the standard error of the weighted mean (N = 7-8, see Methods) and error bars in the fraction of diversity preserved represent the standard error of the weighted mean (N = 8) where weights come from uncertainties in counting the number of sectors.

diversity preserved was calculated as in Ref [26] by dividing the number of neutral sectors by one-half times the estimated initial number of cells at the inoculum front (see Figure 2.1b). The factor of one-half accounts for the probability that two neighboring cells at the inoculum

front share the same color label. The initial number of cells is estimated by measuring the inoculum size of each colony (manually measured by fitting a circle to a brightfield backlight image at the time of inoculation) divided by the effective cell size for *E. coli* ($\sqrt{\text{length} \cdot \text{width}}$, taken to be 1.7 μm, Ref [26]).

Colony fitness

The colony fitness coefficient between two strains was measured using a colony collision assay as described in Refs [26, 37] by growing colonies next to one another and measuring the curvature of the intersecting arc upon collision. Cells were streaked from glycerol stock and a single colony for each strain was inoculated into LB with 120 μg/mL spectinomycin for plasmid retention and incubated at 37°C for 15 hours. The culture was back diluted 1:500 in 1mL fresh LB with 120 μg/mL spectinomycin and grown at 37°C for 4 hours. 1 μL of the culture was then inoculated onto the prepared 12cm×12cm square petri dishes containing LB with different concentrations of chloramphenicol (0μg/mL, 1μg/mL, 2μg/mL, 3 μg/mL) in pairs that were 5 mm apart, with 32 pairs per plate, then the colonies were incubated at 37°C. After half of a day, bright field backlight images are taken and were used to fit circles to each colony to determine the distance between the two colonies. After 6 days, the colonies were imaged with fluorescence microscopy using a Zeiss Axio Zoom.V16. The radius of curvature of the intersecting arc between the two colonies was determined with image segmentation and was used to calculate the fitness coefficient between the two strains (Figure 2.S9a).

Measurement of non-neutral establishment probability

We transformed 9 gene deletion strains from the Keio collection ($\Delta gpmI$, $\Delta recB$, Δpgm, $\Delta tolQ$, $\Delta ychJ$, $\Delta lpcA$, $\Delta dsbA$, $\Delta rfaF$, $\Delta tatB$) and 3 strain backgrounds (BW25113, MG1655, DH5α) with pQY11 (CFP, *specR*) or pQY12 (YFP, *specR*, *cmR*). Cells were streaked from glycerol stock and a single colony of each strain was inoculated into media with 120 μg/mL spectinomycin for plasmid retention, then incubated at 37°C for 16 hours. The culture was back-diluted 1:1000 in 1mL fresh media with 120 μg/mL spectinomycin and grown at 37°C for 4 hours. YFP chloramphenicol-resistant and CFP chloramphenicol-sensitive cells from the same strain background were mixed respectively at approximately 1:500, 1:200, and 1:50 and distributed in a 96-well plate. A floating pin replicator was used to inoculate a 2-3mm droplet from the liquid culture onto prepared OmniTrays with varying concentrations of chloramphenicol (0μg/mL, 1μg/mL, 2μg/mL, 3 μg/mL). Droplets were dried and the plates were incubated at 37°C for 3 days, then imaged by fluorescence microscopy using a Zeiss Axio Zoom.V16.

The establishment probability of the resistant strain can be measured by counting the number of established resistant sectors normalized by the initial number of resistant cells at the inoculum front [26], which gives the probability that any given resistant cell in the

inoculum escaped genetic drift and grew to a large enough size to create a sector. Briefly,

$$p_{est} = \frac{N_{sectors}}{N_0} \tag{2.2}$$

where $N_{sectors}$ is the number of resistant sectors after 3 days (counted by eye) and N_0 is the estimated initial number of cells of the resistant type at the inoculum front. Because the establishment probability can only be accurately measured when the initial number of resistant cells is low enough that the resistant sectors do not interact with one another, we only keep colonies where neighboring resistant sectors are distinguishable at the colony front. In cases where we could see that a sector had coalesced from multiple sectors, we counted the number of sectors pre-coalescence. We also did not find a clear downward bias in the establishment probability as a function of initial mutant fraction (Figure 2.S10), suggesting that the probability of sector coalescence is low in the regime of these experimental parameters. The initial number N_0 of cells of the resistant type is estimated by multiplying the initial number of cells at the inoculum front (see "Measurement of neutral fraction of diversity preserved") by the fraction of resistant cells in the inoculum (measured by plating and counting CFUs).

2.4 Results

Label-free method for measuring demographic noise in microbial colonies

To measure demographic noise in an expanding microbial colony without genetic labels, we developed a method that consists of two steps (Figure 2.1a, Methods, and Supplementary section "Additional information on bead-based sparse lineage tracing method in colonies"). First, we record the trajectories of cell-sized fluorescent beads embedded in the colony (Supplementary movie), which we show track lineages of their neighboring cells, allowing us to track lineages sparsely (i.e. at low-density) (Figures 2.1c, 2.S1c-d, 2.S2, and Supplementary methods). Second, we analyze the fluctuations of the measured lineages (via the bead trajectories) using their length-dependent mean squared displacement (MSD), which serves as an established statistic to quantify the strength of demographic noise by quantifying the randomness in the movement of cells due to growth-induced mechanical forces (Methods and Ref [28]). Intuitively, beads with higher MSDs reflect a colony environment where the mechanical forces being exerted on the beads and cells are more random. Thus, under these conditions the cells that make it to the expansion front where they can more easily reproduce is a more stochastic process and demographic noise is higher.

To determine the ability of our method to measure differences in demographic noise, we compared it to an existing method which uses neutral fluorescent labels to measure the fraction of diversity preserved (the fraction of surviving fluorescent sectors) after a range expansion [28, 27] (Figure 2.1b and Methods). Figure 2.1d shows that the fraction of diversity

preserved is negatively correlated with MSD for a subset of 9 strains (ρ = -0.87, p = 0.002), and can be well fit to an inverse square root relationship (χ_r^2 = 37.5). This inverse square root relationship is consistent with the theory expectation for the fluorescent sector boundary MSD [27], suggesting that the bead MSD captures the fluorescent sector boundary MSD, and is thus a convenient and reliable measure of demographic noise. We chose to report the MSD at the window length L = 50μm because the inverse square root fit to the fraction of diversity preserved had the lowest chi-squared at this length scale (Figure 2.S7). To control for growth rate differences between the strains, we also masked the colonies with the smallest colony's outline and remeasured the fraction of diversity of preserved; this did not significantly change the ordering of the genotypes (Figure 2.S1b).

The distribution of demographic noise for single gene deletions

We next wanted to use our bead-based sparse lineage tracing method to measure the distribution of demographic noise due to single gene deletion mutations. We randomly selected 191 single gene deletion strains from the Keio collection [5], a well-characterized library of *E. coli* strains that contains all non-lethal single gene deletions (see Methods). In order to test such a large number of colonies, we grew the colonies in 96-array format on multiple agar plates. We observed variation in bead MSD between plates (see Supplementary information "Sources of variation"), and therefore report the bead MSD of the gene deletion strains relative to that of the wild type, which is present in 8 replicate colonies per plate (see Methods).

Figure 2.2 shows the distribution of relative MSD from the 191 randomly selected single gene deletion strains. The knockout (KO) distribution significantly differs from the wild type (WT) distribution (Kolmogorov-Smirnov p = 2.7×10^{-4}) with a lower mean (KO: 0.904, WT: 1.011) and higher variance (KO: 0.044, WT: 0.004). 39% of knockout MSDs were lower than the lowest wild type MSD observed, with the maximal decrease in knockout MSD of 81% from the wild type median MSD. Interestingly, the typical knockout mutation decreases demographic noise from that of the wild type by 8% (95% CI = [1%, 17%], Supplementary Methods) (Median KO = 0.94, Median WT = 1.02).

To determine whether any biological processes or pathways could explain the differences in demographic noise observed, we performed a gene enrichment analysis by looking for GO and KEGG terms whose average MSD value across gene knockouts was significantly different from (1) that of randomly selected knockouts (i.e. what causes 81% vs 8% decrease in genetic drift?) and (2) that of the wild type (i.e. what causes a different strength of genetic drift in KOs vs WT?) (Supplementary Methods). For the first analysis, there were no significant GO terms identified and only a single significant KEGG term (ATP-binding cassette reporters) (see Supplementary information "Gene enrichment analysis"). For the second analysis, 6 out of 8 KEGG terms represented in the randomly selected subset of knockouts from the Keio collection had significantly lower average MSDs from the wild type (Figure 2.S11). Many of the significant KEGG terms relate to metabolism (carbon metabolism, metabolic pathways, microbial metabolism in diverse environments, biosynthesis of secondary metabolites). As

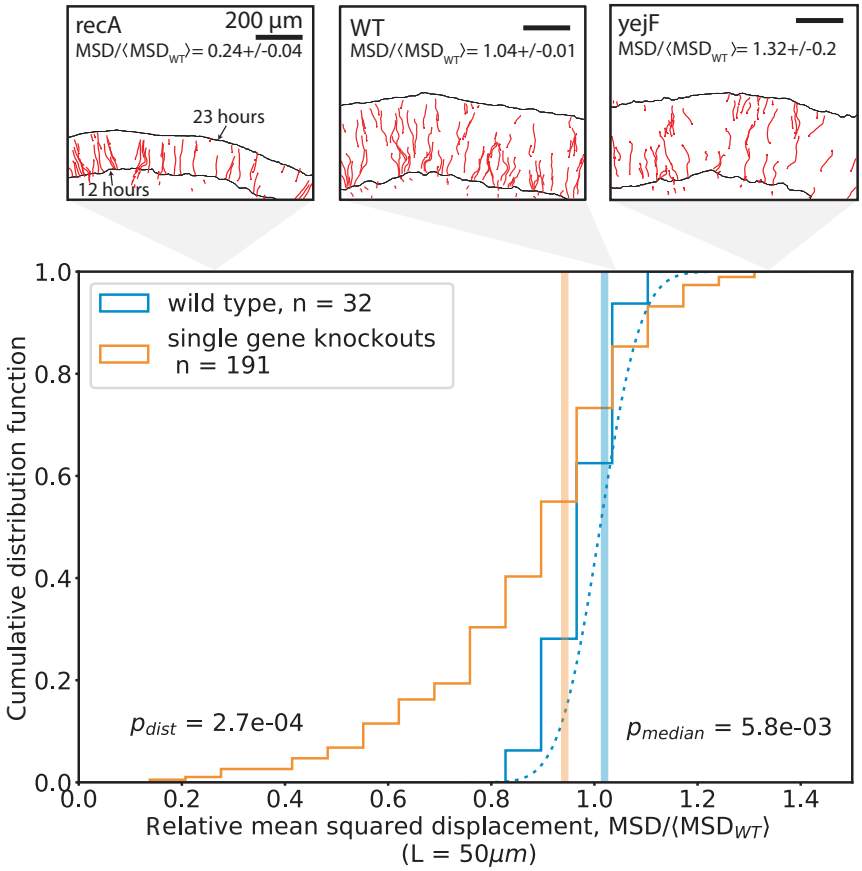

Figure 2.2: Distribution of MSD across randomly-sampled Keio collection single gene deletion strains. Each count in the distributions is an average of 2-3 replicate colonies grown on different agar plates, and the MSD is normalized to the average wild type MSD on each plate (Methods). The blue dotted line shows a Gaussian fit to the wild type distribution. Vertical lines show the median value of each distribution. Panels show examples of bead trajectories from wild type and single gene deletions strains from the Keio collection. Black lines in panels show the colony front at t = 12 hours and t = 23 hours.

metabolism relates to growth rate, this suggests that growth rate plays a role in determining the strength of demographic noise, which is consistent with our finding described in the next section that colony area (a proxy for biomass) positively correlates with demographic noise. Many significant GO terms were identified in the second analysis, possibly because the knockout and wild type distributions are significantly different from one another, making them hard to interpret.

Phenotypic trait predictors of the strength of demographic noise

We noticed that some colonies with particularly low bead MSD also seemed to be small with smooth colony shapes. As a result, we systematically checked which phenotypic traits best correlate with the observed differences in strengths of demographic noise in the single gene deletions. Specifically, we measured a range of traits in the same colonies for which we measured bead MSD (see Methods), including the depth of the growing layer of cells at the front, the roughness of the colony front, the area of colony, and we also used existing datasets for single cell shape. While previous studies have studied the relationship of these traits with demographic noise experimentally by comparing species or strains in low-throughput [26, 14, 58], the same strain in different nutrient concentrations [46], or using simulations [49, 14], our system allows us to experimentally test correlations of demographic noise with different traits in a large number of related strains, thus overcoming a major experimental limitation.

We found that (1) the roughness of the colony front is positively correlated with the bead MSD (Figure 2.3a, Pearson r = 0.66, p = 2×10^{-23}), (2) the size of the growing layer of cells at the front of the colony is negatively correlated with the bead MSD (Figure 2.3b, Pearson r = -0.54, p = 4×10^{-15}), and (3) the colony area after 1 day of growth, which we checked can be used as a proxy for biomass (see Supplementary methods and Figure 2.S12a-b), is positively correlated with the MSD (Figure 2.3c, Pearson's r = 0.63, p = 5×10^{-21}). These colony-level results agree with theoretical predictions [49, 14] and previous experimental results [46, 14]. Using datasets of single cell shapes from the Keio collection from Refs [61, 17], we did not find a significant correlation of demographic noise with cell shape (Figure 2.S13a-b), in contrast to the colony-level traits.

We estimated the joint relationship of the measured traits with bead MSD using Lasso regression [60], which finds the minimal set of traits that predict the MSD and the coefficients associated with those traits in a linear model. The traits that were included in the Lasso regression were the 4 colony traits (front roughness, growth layer depth, colony area, and colony thickness) that we measured and the 5 single-cell shape traits (aspect ratio, minor axis length, surface area, volume, and major axis length) from the dataset in Ref [17]. We find that all 4 colony level traits and single cell aspect ratio are the only 5 traits included in the best fit model to the MSD (Figure 2.3e), with the coefficient for the single cell aspect ratio being almost an order of magnitude lower than that of the lowest colony trait coefficient. Using the best fit model, we are able to explain the variance in the MSD with an R^2 of 62% (Figure 2.3f).

These correlations of demographic noise with various population-level traits could be partly driven by correlations between the population-level traits themselves. Indeed, we find that colonies with larger areas tend to have smaller growth layer depths (Figure 2.S14b) and higher front roughness (Figures 2.S14a). Prior theoretical studies have suggested that colony traits are interdependent [49, 14]: faster growing strains have a sharper nutrient gradient at the front, leading to a smaller growth layer, which in turn creates more front roughness, which is consistent with our findings. Both the correlation of colony area with front roughness and that of colony area with bead MSD across strains could potentially be explained without

demographic noise differences if the front roughness and bead MSD increased over time within a single colony as it grew larger. In order to exclude this possibility, we checked that the front roughness saturates over time by the time of measurement (Figure 2.S12f) and that the bead MSD does not increase as the colony grows larger but rather slightly decreases (Figure 2.S12e). In order to test which traits can be a causal determinant of MSD, we corrected for linked correlations between traits using partial correlations (Figure 2.S15), and find a slightly lower but significantly nonzero correlation of bead MSD with front roughness ($r = 0.53$, $p = 2 \times 10^{-13}$) and a more substantial decrease in correlation of bead MSD with colony area ($r = 0.29$, $p = 2\times10^{-4}$) and growth layer depth ($p = -0.17$, $p = 2.5\times10^{-2}$). This supports the idea that front roughness is the main causal determinant of MSD, as was also shown in Ref [14].

Because previous work has found that colonies grown from cells with round shapes tend to have lower demographic noise [26, 58], we were puzzled by our result showing lack of correlation between cell shape and bead MSD. Thus, we specifically tested a round cell shape mutant, MC1000 $\Delta mreB$, which we indeed measured to have low MSD compared to the wild type MC1000 (Figure 2.S13e). However, by using the best fit Lasso regression model (Figure 2.3e), which primarily includes colony-level traits, the low MSD could be predicted (Figure 2.S13e), suggesting that colony-level traits are sufficient to explain the difference in MSD. Because it is possible that differences in colony-level traits mask the effect of single-cell traits on bead MSD, we also corrected for variation in all other traits using partial correlations; however, the corrected bead MSD still shows little correlation between cell shape and bead MSD (Figure 2.S13f). We note that we cannot rule out the possibility that the lack of correlation between bead MSD and cell shape for the Keio mutants described above using the datasets from Refs [17, 57] is influenced by differences in cell shape exhibited by cells of the same genotype in different growth conditions (see Figure 2.S13c-d and Supplementary section "Comparison of cell shape between growth in liquid culture and as a colony"). We also measured a library of *mreB* and *mrdA* point mutants that were enriched for cell shape differences (45). In this enriched library, there is a slightly higher correlation between cell shape and the strength of demographic noise (Pearson $r = 0.32$ and $r = 0.46$ for *mreB* and *mrdA* point mutants respectively, see Figure 2.S13g-h), possibly because the cell shapes span a larger range, or because our growth condition was more similar to that of the single cell measurements in this dataset.

In summary, Lasso regression suggests that a combination of colony-level traits best predicts bead MSD, which we have shown is anticorrelated with demographic noise. However, after correcting for correlations between phenotypic traits, we found evidence supporting that the main causal determinant of MSD is the colony front roughness. Additionally, the agreement of the colony-level phenotypic trait relationships with those found in previous work [49, 26, 46, 14] suggests that the same mechanisms for how phenotypes affect demographic noise seem to hold in range-expanding populations regardless of whether looking across single mutations, different strains, or different species.

Single gene deletions can substantially alter adaptation through changes to demographic noise

Finally, we sought to determine whether the variation in demographic noise induced by single deletions also induces a substantial corresponding change in evolutionary outcomes, such as the establishment probability of beneficial mutations, as predicted by population genetics theory. We constructed fluorescently-labeled chloramphenicol resistant and sensitive strains on selected strain backgrounds and measured the establishment probability of the resistant type when competed with the sensitive type on the same strain background (Figure 2.4a and Methods). The fitness coefficient between the resistant and sensitive types was tuned by the chloramphenicol concentration and measured for each strain background at each chloramphenicol concentration using a colony collision assay (Methods and Figure 2.S9a).

We found that the establishment probability of the resistant type is negatively correlated with the bead MSD of the strain background across beneficial fitness coefficients from s = 0.05 to s = 0.15 (Figure 2.4b). The Keio collection WT had the largest MSD and lowest establishment probability of the Keio collection strains that were tested. The maximal increase in the establishment probability of a beneficial mutant on a gene deletion background was about 6-fold over the WT, which corresponded to about a 4-fold decrease in the background strain MSD compared to the WT. Interestingly, we observed that changing the initial fraction of the resistant type sometimes changed the establishment probability (Figure 2.S16), possibly due to interactions between beneficial sectors; however, we did not detect any systematic effect across strain backgrounds (Figure 2.S10). We controlled for differences in initial fraction by separating the data by initial fraction, and found that the effect of the initial fraction of the resistant type on the establishment probability does not explain the observed negative correlation between demographic noise and establishment probability (see Figure 2.S17b). In sum, we find that the range of strengths of demographic noise accessible by single gene deletion strains substantially affects the establishment probability of a beneficial mutant on that background.

2.5 Discussion

We have shown that single gene deletions can substantially alter the strength of demographic noise in microbial colonies (Figure 2.2) and that these differences can have an impact on adaptation (Figure 2.4). We accomplished this by developing a bead-based sparse lineage tracing method for measuring demographic noise in colonies (Figure 2.1). While the beads could potentially perturb the cell lineages, possibly impacting the correlation of bead MSD with the fraction of diversity preserved, both quantities allowed us to observe sufficient differences between strains. We checked whether there were particular types of genes that altered demographic noise and found that genes associated with KEGG terms relating to metabolism were enriched for lower strength of genetic drift (Figure 2.S11). We additionally used this method to measure a non-random set of strains from the Keio collection as well

as mreB and mrdA point mutants and found an even larger range of demographic noise effects (Supplementary section "Additional measurements of distribution of drift effects" and Figure 2.S18).

Our results suggest that demographic noise itself may be an evolvable trait of a population. We hypothesize that strain backgrounds with different strengths of demographic noise may also exhibit different rates of adaptation when accumulating multiple mutations. It would be interesting to test this hypothesis in future work empirically through experimental evolution of colonies [9, 62] and theoretically through simulations with joint distributions of demographic noise effects and fitness effects. Quantitatively, the evolution of demographic noise may be similar to the evolution of mutation rate, because both mutations and demographic noise primarily influence the establishment rate of new mutations; however, this should be examined more carefully in future work in different regimes such as successive mutations and clonal interference [24]. Additionally, interesting dynamics could arise in spatially-structured communities with cooperation, such as those that share a common good [49]. Increasing demographic noise in these systems may make cheating less likely by leading to more spatial segregation of cheater and producer types. Another interesting corollary to our results is that a decrease in the strength of demographic noise enables more efficient transfer of genetic material through conjugation in bacterial colonies [16] and exchange of metabolites between co-expanding strains [23, 48].

Our results show that demographic noise is correlated with colony-level traits (Figure 2.3), suggesting that the strength of demographic noise in these colonies is set by collective behavior. As a result, we hypothesize that the plasticity of demographic noise holds more generally in self-organized systems [33], including colonies, biofilms, spatially-structured microbiomes, and solid cancer tumors, which would be interesting avenues for future study. Additionally, other phenotypic traits have been predicted to influence colony patterning and demographic noise and it would be interesting to test their influence on demographic noise in future work, including that of cell-cell and cell-substrate adhesion [22, 32, 52], cell orientations [14], cell elasticity [14], and variation in single cell growth rates [41, 56] and lag times [64, 40].

The positive correlation between colony area and demographic noise (Figure 2.3c, $r = 0.63$) suggests a tradeoff between demographic noise and fitness (Figure 2.S19): a beneficial mutation may increase demographic noise and actually impair its own establishment and once established, also the establishment of future beneficial mutations. However, when a demographic-noise-modifying mutant first arises at a low frequency in a colony, the colony-level traits will be set by both that of the mutant and the background strain, so the strength of demographic noise that governs its trajectory will likely be a complex time-dependent combination of traits from the two genotypes in monoculture. Thus, while this work generates interesting hypotheses as to the tradeoffs between demographic noise and fitness, future work is needed to more closely examine the consequences of demographic noise and fitness correlations in different environments.

The bead-based sparse lineage tracing method in colonies can be extended to study demographic noise in other genotypes in high-throughput, such as double-mutants and potentially other species. In the supplementary text, we use this method to measure demographic

noise in *S. cerevisiae* colonies (Figure 2.S1e), and we find a lower bead MSD in *S. cerevisiae* compared with that of *E. coli*, in agreement with results from previous work [26, 28]. Measurements of demographic noise in additional genotypes can be used to understand the dependence of the distribution of strength of demographic noise on the genetic background across large mutational differences (different species) or small mutational differences (double-mutants).

Like selection, mutation, migration, and recombination, demographic noise has been shown to be an important evolutionary force in many systems. Understanding the environmental and genetic influences on demographic noise will allow us to better identify and model the relevant forces that drive evolution in different systems. Whereas demographic noise is typically thought of as being static or dependent on the environment, we have shown that like for other evolutionary forces, demographic noise can be considered an evolvable trait of a population. Future work exploring the evolvability of demographic noise will help us better understand its consequences on evolutionary outcomes in different systems.

Data availability

Data and code are available at `github.com/qinqin-yu/colony-demographic-noise`.

Acknowledgements

We thank Jona Kayser, Joao Ascensao, Benjamin Good, and the Hallatschek lab for valuable discussions and feedback. We thank KC Huang and Handuo Shi for sharing the *mreB* and *mrdA* point mutation libraries, and Christophe Beloin for sharing the quandruple adhesin and polysaccharide mutant strains. We thank the Laura Waller lab for lending the objective used for cell shape validation. We thank two anonymous reviewers for very constructive comments on the final manuscript. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award R01GM115851 to OH, a National Science Foundation CAREER Award (#1555330) to OH and a Simons Investigator award from the Simons Foundation (#327934) to OH, a National Science Foundation Graduate Research Fellowship under Grant No. (DGE 1106400) to QY, and a fellowship from the Simons Foundation's Society of Fellows (#633313) to AH.

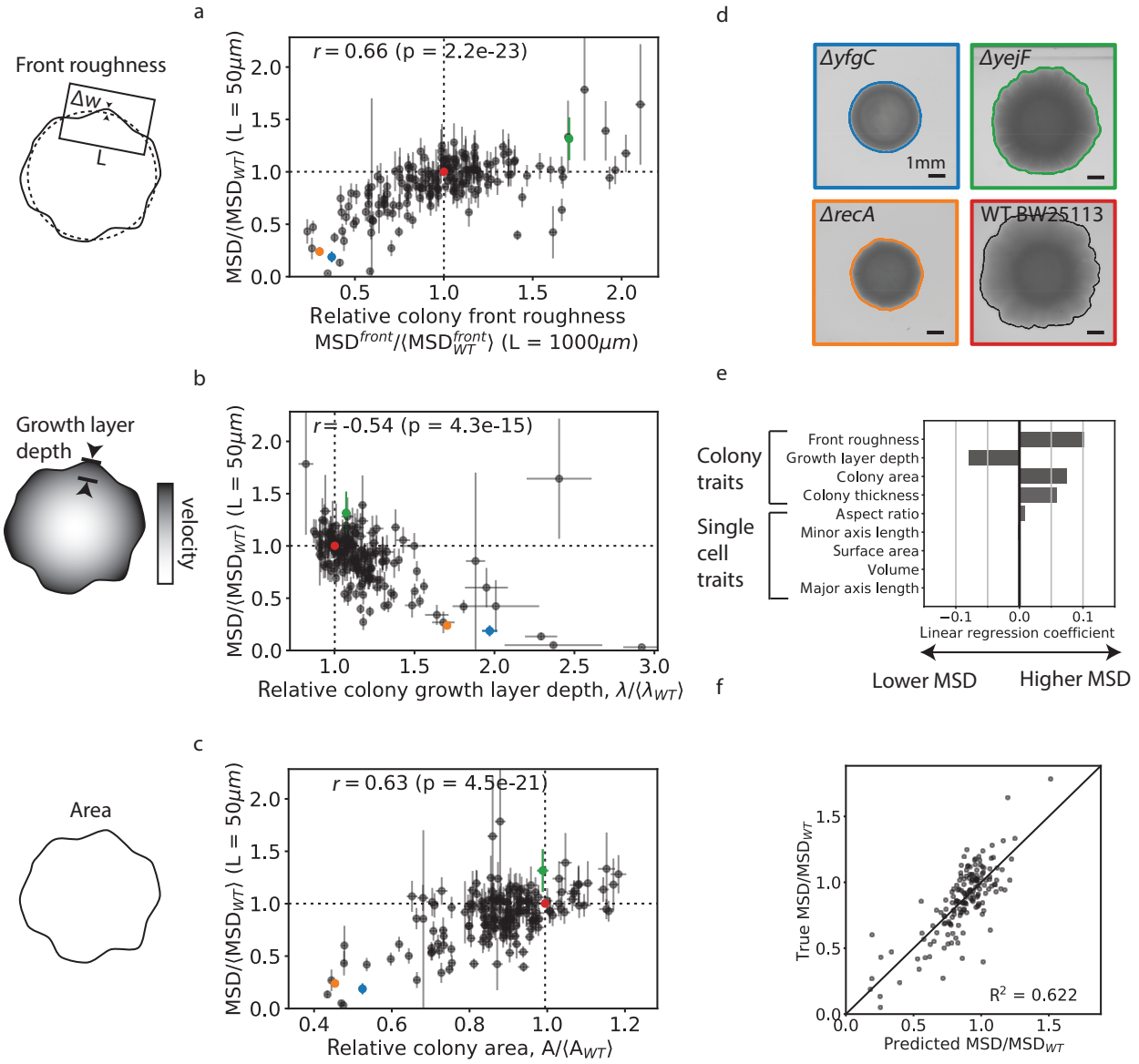

Figure 2.3: Phenotypic predictors of the strength of demographic noise. Correlation of the bead trajectory MSD for 191 single gene deletions and 41 selected strains with (a) front roughness (defined in Supplementary methods), (b) colony growth layer depth (defined in Supplementary methods and Figure 2.S12c-d), and (c) colony area. Error bars represent the standard error of the mean across 2-3 replicate colonies. (d) Example colonies for colored points in (a)-(c) (e) Linear model coefficients for phenotypic traits that best predict MSD, estimated through Lasso regression. (f) Predicted MSD using the linear model with the coefficients shown in (e).

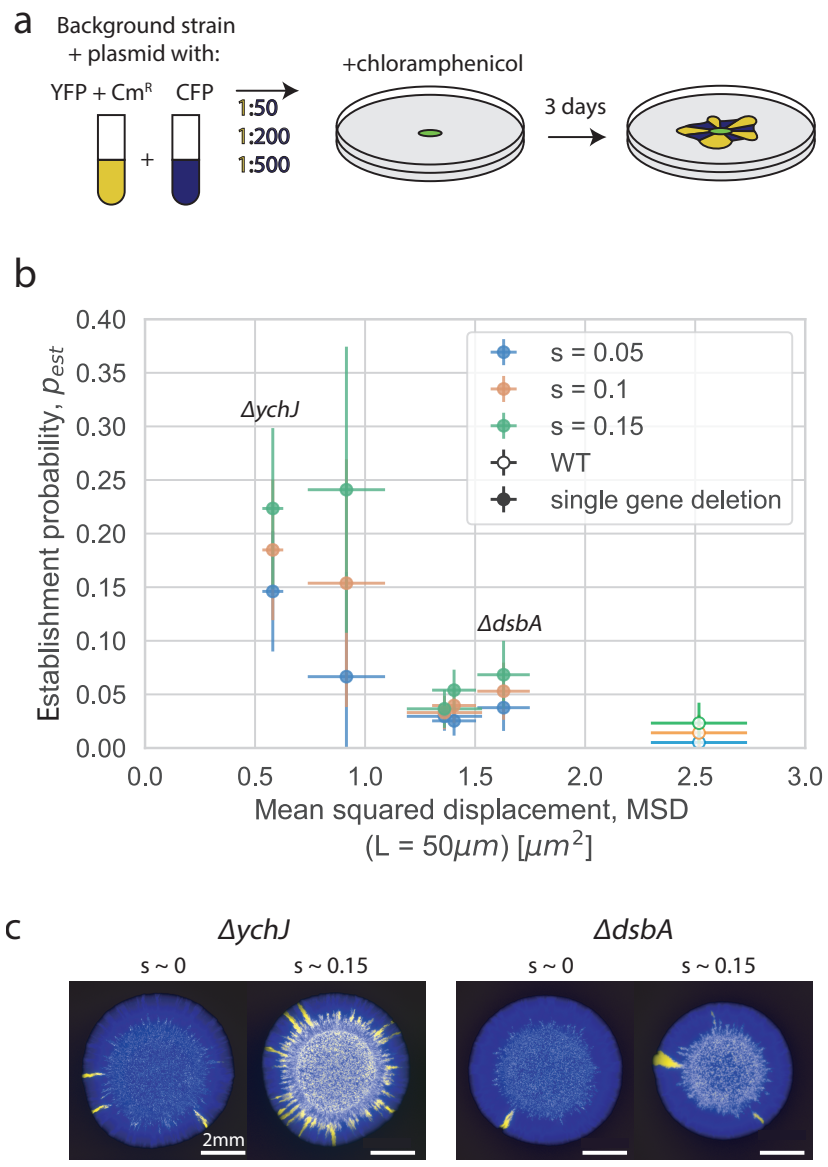

Figure 2.4: Single gene deletions can substantially alter adaptation through changes to demographic noise (a) Schematic for measurement of establishment probability. After 3 days, the number of beneficial sectors is counted and the establishment probability is calculated by dividing by the initial number of chloramphenicol resistant cells at the expansion front [26]. Cm^R: chloramphenicol resistance gene. (b) Interpolated establishment probability at three different fitness coefficients as a function of bead trajectory MSD. Error bars in establishment probability represent linear fitting error (see Supplementary methods) and error bars in bead MSD represent the standard error of the weighted mean (N = 7-8, see Methods) (c) Example colonies for two different strain backgrounds each at two fitness coefficients between the resistant and sensitive types.

2.6 Supplementary methods

Strains

Strain	Description	Genotype	Derived from	Antibiotics	Ref
BW25113	Keio collection background		*E. coli* K-12		[5]
BW25113 single gene deletions	Keio collection	see Supplementary Table 1	BW25113		[5]
MG1655			*E. coli* K-12		
DH5α					
MC1000					[38]
MC1000 $\Delta mreB$			MC1000		[38]
$\Delta_4$pol	Deletion mutant of polysaccharides Yjb, cellulose, PGA and colonic acid	MG1655 $\Delta yjbEH$ :: cm $\Delta bcsA$ :: KmFRT $pgaA$ :: $uidA - zeo$ $cps5$ ::Tn10			[12]
$\Delta_4$adh	Deletion mutant of flagella, AG43, type 1 fimbriae, and curli	MG1655 gfp $\Delta fliER$ :: cm $\Delta fimAH$ :: zeo Δflu ::FRT $\Delta csgA$ :: $spec$			[15]
RDM 893	$mreB$ point mutant background	MG1655 $\Delta mreB$ pRMmreBCD			[57]
TKL 117	$mrdA$ point mutant background	MG1655 $\Delta mrdA$ pRMindpbp2			[57]
$mreB$ point mutants		see Supplementary Table 2	RDM 893	15 μg/mL chloramphenicol	[57]

mrdA point mutants		see Supplementary Table 2	TKL 117	35 μg/mL kanamycin and 50 μM IPTG	[57]

Plasmids

Plasmid	Description	Genotype	Derived from	Antibiotics	Ref
pQY10		Venus YFP (A 206K), $Spec^R$		120 μg/mL spectinomycin	this study
pQY11		(e)CFP (A 206K) , $Spec^R$		120 μg/mL spectinomycin	this study
pQY12		Venus YFP (A 206K), $Spec^R$, Cm^R		120 μg/mL spectinomycin, variable chlorampenicol	this study
pQY13		(e)CFP (A 206K) , $Spec^R$, Cm^R		120 μg/mL spectinomycin, variable chlorampenicol	this study

Creation of glycerol stocks for measurement of distribution of demographic noise

We created master stocks by rearraying strains from the original frozen Keio collection glycerol stocks into 96-well plates with LB, growing overnight, and freezing the rearrayed cultures in a 25% glycerol stock. Master stocks of the mreB and mrdA point mutant libraries [57] were acquired from the lab of KC Huang. We created separate glycerol stocks that could be defrosted for each experiment by scraping frozen glycerol stock from the master stocks into a 96 deep well plate filled with 1 mL LB with the appropriate antibiotics, growing overnight until saturation, aliquoting 100 μL into multiple 96-well plates, and freezing as separate glycerol stocks.

Single particle tracking

Single particle tracking was performed using a custom-written code in MATLAB based on that used in Ref [12]. Briefly, particle image velocimetry (PIV) was first used to subtract large-scale movements between frames. Then, particles with a radius smaller than 3 pixels were detected and linked between consecutive frames with single particle tracking to sub-pixel accuracy. Displacements between consecutive timepoints were then joined to create trajectories spanning multiple timepoints. The trajectories of particles in regions without cells were used to calibrate xy fluctuations in stage position over time. After correcting for stage fluctuations, trajectories that were shorter than 30 μm in total length were rejected as stationary beads and consecutive timepoint steps that were shorter than 1 pixel were rejected as noise.

Measurement of colony front roughness

The front roughness was measured using the method described in Ref [26]. A backlight image of the colony was taken at 24 hours, and a custom-written MATLAB code was used to extract the colony boundary using image segmentation. The boundary was fit with a circle and the mean squared displacement in the radial direction was calculated for windows of different arc lengths ($L = 200$ linearly spaced arc lengths from 6 to 1152 μm) along the best fit circle using a running average over overlapping definitions of the starting position of the window. The MSD was fit to a power law as a function of L and the value of the fit at $L = 1000\mu$m was reported.

Measurement of growth layer depth

The growth layer depth was measured using bead displacements between pairs of consecutive timepoints (Figure 2.S12c). We assumed that the velocity v along the direction of growth has the form

$$v_y(y) = v_{y,max} e^{-(y-y_0)/\lambda} \tag{2.S1}$$

where $v_{y,max}$ is the maximum velocity, y is the position along the direction of growth, y_0 is the front position, and λ is the growth layer depth. The ratio v at two positions within the colony then conveniently takes on the form

$$v = \frac{v_y(y_2)}{v_y(y_1)} = e^{-(y_2-y_1)/\lambda}. \tag{2.S2}$$

The average direction of motion of all the trajectories in a field of view was determined, and all subsequent measurements were projected along this axis. For each pair of consecutive timepoints, and for each pair of beads, the ratio of their displacements, $\frac{v_y(y_2)}{v_y(y_1)}$, and the distance between them parallel to the average direction of motion, $\Delta y = y_2 - y_1$, were calculated. Because noise from inaccurate tracking can dominate the ratio of the velocity

when one or both of the bead velocities is small, we restricted our analysis to $-50\mu m < \Delta y < 50\mu m$. To minimize the effect of changes in curvature of the front of the colony, we restricted pairs of beads to have $\Delta x < 50\mu m$. We binned the data (bin size = $10\mu m$, or if range of $\Delta y < 100\mu m$ then we used a bin size of std(Δy)/2) and fit to an exponential decay function using weighted least squares to extract the growth layer depth λ (Figure 2.S12d). The error in λ represents the error in the fitting parameter.

Measurement of colony thickness

The thickness of the colony was determined using a backlight brightfield image of the whole colony at 23h, where the exposure time was fixed for all colonies. We calibrated the intensity of the transmitted light to the colony thickness using a colony that was grown with fluorescent tracer beads (Methods in main text). We noticed that beads primarily rise to the top of the colony, so we used the height of the fluorescent beads as the ground truth colony height. To measure the average thickness of the whole colony, we calculated the average brightfield intensity within each colony and converted it into a thickness using the calibration curve. We note that this method does not distinguish between changes to intensity due to changes in colony thickness and changes in biomass density, but we make the assumption that any changes to density are small in comparison to changes in thickness.

Single cell shape comparison in liquid and colony conditions

Cell preparation for growth in liquid culture was closely matched to the protocols of the National BioResource Project described in Refs [61] and [17]. Cultures were grown overnight with rotation at 30C, back-diluted 1:100 in fresh LB medium the next day and then incubated at 37C for 2 hours. Before imaging, cells were diluted in PBS to achieve the desired imaging density and vortexed to break up clumps. For imaging, a droplet of the diluted culture was placed on an agar pad (LB + 2% agar), and covered with a cover slip to prevent evaporation.

Cell preparation for growth in colonies was closely matched to the protocol for growing colonies to measure demographic noise in this work. Cultures were grown overnight with rotation at 37C. The next day, a 2μL droplet of the saturated culture was placed on a plate with LB and 2% agar. The colony was grown at 37C for 1.5 days. Cells were picked from the edge of the colony and resuspended in LB, and vortexed to break up clumps. For imaging, a droplet of the resuspended culture was placed on an agar pad (LB + 2% agar), and covered with a cover slip to prevent evaporation.

Cells were imaged with phase contrast microscopy using a Nikon Eclipse Ti-E inverted microscope with a 40x, 0.65 NA phase contrast air objective (Nikon, Düsseldorf, Germany). Image segmentation was done with the software Morphometrics [61] and features that were incorrectly segmented were manually rejected. The circularity C is defined as

$$C = \frac{l^2}{4\pi A} \tag{2.S3}$$

where l is the cell contour length, and A is the cell area. The circularity of a circle is 1 and values larger than 1 correspond to more elongated cells.

Single cell imaging with beads

In order to track the movement of the cells and beads at the single-cell resolution (Figure 2.1c and Figure 2.S1c), we grew a colony following the procedure in the main experiments for demographic noise for 18-24 hours at 37C, then covered the colony with a BSA-coated coverslip. The sample was incubated at 32C with a H201-T Okolab incubator (Ottaviano, Italy) and a home-built incubation chamber for imaging. We imaged the cells and beads with an Olympus IX81 Inverted microscope with a 40x 1.3 NA oil objective (Olympus, Hamburg, Germany) in brightfield. Images were taken every 2-4 minutes for approximately 1 hour. To correct for focus drift due to evaporation, we used μManager to move the z position of the stage at a constant rate and we also took a z stack of eleven 1μm slices which were postprocessed to find the most in focus image. The beads could be identified from their different scattering properties compared to the cells.

Gene enrichment analysis

GO and KEGG terms associated with at least 3 genes in the 191 randomly chosen strains were tested for the hypothesis that the weighted median MSD value of the knockouts associated with a given term was more extreme than that of a randomly chosen subset of the same number of genes from all knockouts or the wild type distribution. For each term, random subsets of MSD values were drawn 10^5 times when comparing to knockouts and 10^6 times when comparing to the wild type. The fraction of times that the weighted median of the random subset was more extreme than that of the true values was recorded and the p-value was calculated as twice this fraction (due to only looking at the distribution on one side of the median). A Benjamini-Hochberg FDR correction [6] was applied to account for multiple testing, and terms that were significant at a 5% level were kept.

Lasso regression

We used the Lasso regression implementation in the Python library scikit-learn [53] to determine the minimal set of traits that predict the MSD and the coefficients associated with those traits in a linear model. The data from each input phenotype is standardized to remove the mean and scale to unit variance. The cost for the regularization term is determined using 10-fold cross validation to be $\alpha = 5 \times 10^{-4}$.

Partial correlation

The partial correlation was determined by first calculating the linear least squares regression coefficients between the trait of interest and all other traits, and between the bead MSD and

all other traits except the trait of interest. Where errors for the trait of interest existed, they were used to weight the least squares fitting with a weight of $1/err^2$. The residuals were calculated for the trait of interest and the bead MSD. The partial correlation was calculated as the correlation between the residuals.

Comparing establishment probability to bead MSD (Figure 2.4b)

For each genotype, at a particular chloramphenicol concentration, the replicate fitness (N=8) and establishment probability measurements (N=24) were randomly paired. A linear fit was performed using numpy.polyfit on the establishment probability as a function of fitness across all chloramphenicol concentrations and all initial mutant fractions for the fitness coefficient range from s = -0.1 to s = 0.5 since the colony collision assay is only valid for small fitness coefficients [37] (see Figure 2.S16). The establishment probability was calculated using the fit parameters at the fitness coefficients s = 0.05, 0.1, and 0.15. The error in the establishment probability was taken as half of the difference between the maximum and minimum values possible using combinations of fitted slope and intercept parameters that are one standard deviation away. We note that while the establishment probability and fitness are not expected to depend linearly on one another according to theory [39], the linear relationship is a convenient approximation for our data where we don't have sufficient signal to distinguish between more complex models.

Statistical methods

For calculating the confidence interval on the percent difference in medians between the knockout and wild type demographic noise distribution, a modified block jackknife method was used. 10^5 subsets of 32 KO MSD values (the same number as in the WT distribution) were drawn (without replacement in each subset) and their medians were calculated. The confidence interval was taken from the 2.5% to the 97.5% of the empirically measured subsampled median distribution.

2.7 Supplemental information

Additional information on bead-based sparse lineage tracing method in colonies

Demographic noise in colonies has previously been quantified by measuring the number or shape of sectors in whole colonies [28] or tracking single cell lineage dynamics in a growing microcolony [26]. However, these methods require fluorescently labeled strains or time-intensive imaging and analysis which makes them impractical for screening large numbers of strains.

Our label-free method to measure demographic noise in microbial colonies allows us to screen a large number of strains and to minimize the probability of accumulating new mutations by avoiding genetic transformation. This is important as we aim to test the effect of single loss of function mutations on demographic noise, and additional mutations could potentially change the strength of demographic noise that is measured. The method also allows potential future study of microbes that are not genetically tractable where genetic transformations with fluorescent markers would not be possible.

As described in the main text and shown in Figure 2.1a, we use spherical fluorescent polystyrene tracer beads to sparsely track cell lineages in colonies and we use the wandering statistics of the bead trajectories to infer the strength of demographic noise. Because the beads are at a lower spatial density than cells, we are able to image with an air objective with a lower NA and a lower magnification than if we imaged single cells, and we can also capture a larger field of view. We can also track the beads with images that are taken less frequently because beads move fewer pixels at a lower magnification, and this allows us to image many colonies in parallel. We note that we tried beads with different coatings (unmodified functional group, aminated functional group, Concanavalin A-coated) and the beads with unmodified functional groups yielded the longest bead trajectories. We interpreted this result as the unmodified beads were the best at following the cell trajectories rather than getting lost behind the front. Thus, we chose to use unmodified beads for our experiments.

As a first approach to validate the method, we compared the measurement of MSD from sector boundaries, single cell trajectories, and bead trajectories for *E. coli* DH5α and *S. cerevisiae* W303 (Figure 2.S1e), which are two species whose strengths of demographic have previously been compared [26, 28]. In all three methods, *E. coli* has a higher MSD than *S. cerevisiae* as previously measured [26, 28]. The three methods span different length scales, with the single cell lineages spanning the shortest length scale, the bead tracks spanning an intermediate length scale, and the sector boundaries spanning the longest length scale. There is a continuous transition from the MSD of *E. coli* single cells to beads; however, there is a discontinuity between the transition from bead trajectories to sector boundaries. Additionally, there are discontinuities in the transition between the three methods for *S. cerevisiae*. The discontinuity between the bead trajectories and the sector boundaries may result from movement of the sector boundaries behind the colony front due to continued growth and realignment of cells. Another source of the discontinuities may come from the

finite spatial resolution for the sector and bead images.

As a second approach to validate the method, we imaged the beads and single cells for the wild type strain BW25113 in the same field of view to determine if the bead trajectories were following the cell lineages. Visually, we see that the bead trajectories follow at least 1 cell lineage (traced backwards in time) at the colony front and behind the front (Figure 2.1c and Figure 2.S1c). We unfortunately were not able to image for a longer than 1 hour because the images went out of focus. Additionally, we note that adding a coverslip on top of the colony, which was necessary for using a high NA oil immersion objective to get single cell resolution, may change the mechanical behavior of the cells.

As a third approach to validate that the method can measure demographic noise, we compared the bead trajectory MSD to the fraction of diversity preserved in a neutral fluorescent mixture after 1 day of growth (Figure 2.1d). As discussed in the main text, we find fitting an inverse square root relationship to the fraction of diversity preserved as a function of the MSD roughly matches the shape of the relationship. We emphasize that our goal is not get the most accurate measurement of demographic noise, but to screen a large number of strains to determine whether there is a significantly different distribution of demographic noise effects from single gene knockouts compared to the wild type. Thus, small deviations from the fit can still allow us to measure the distribution of demographic noise as long as the deviations are less than the width of the measured distribution.

As beads can fall behind the front, the bead trajectories combine the dynamics at the front and behind the front. Since prior work has shown that successful lineages only come from the first layer of cells at the colony front [26], we tested how closely the full bead trajectories matched the bead trajectory only while it was at the front. We separately measured the MSD of trajectories at the front (within 6μm of the front) and behind the front. Figure 2.S6 shows for wild type BW25113 that trajectories at the front exhibit a similar MSD to the full trajectories, whereas trajectories behind the front exhibit a lower MSD. Thus, the MSD of the full trajectories captures the MSD of trajectories at the front.

Determining the mean squared displacement window size

In order to be able to compare a single number for the MSD for different genotypes, we reported a summary statistic of the MSD at a window size of $L = 50$ μm which is interpolated or extrapolated from the power law fit to the MSD across all window sizes. This window size was chosen because it gave the best fit of fraction of diversity preserved (y) to MSD (x) to the relationship $y = a\sqrt{x}$ [27] amongst the window sizes 25 μm, 50 μm, 100 μm, and 1000 μm (see Figure 2.S7). The fit at 25 μm, 50 μm, and 100 μm gave similar reduced chi-squared values, while the fit at 1000 μm had a much higher reduced chi-squared. The poor fit at high window sizes is likely due to the fact that few bead trajectories become that long, and the extrapolation becomes noisy. Note that while the reduced chi-squared values are generally high, we were not able to account for the errors in MSD in the reduced chi-squared calculation; taking those errors into account should reduce the discrepancies in the summary statistic describing the deviation between the fit and the data.

We also show the distribution demographic noise for the wild type and knockout strains using the MSD reported at different window sizes in Figure 2.S8. At all window sizes except the largest window size of $L = 1000\mu$m (where the MSD also did not agree as well with the fraction of diversity preserved) we see that the wild type and knockout distributions are significantly different from one another to at least a level of 5% (Kolmogorov-Smirnov test).

Sources of variation

We tested for variation across positions on a plate and across different plates. While all media for the experiments described in the main text (except single cell tracking experiments) were made on the same day, variation between plates can result from: the exact ratios of LB powder, agar, and water, the temperature of the liquid as it is poured into a plate, the tilt of the plate while drying, the age of the plate at the time that it is used, and other hidden experimental parameters. Figure 2.S3a shows that the Pearson correlation coefficient for the same strain grown in the same position on different plates was 0.4-0.82 for randomly sampled knockout strains (DE1-4), 0.65-0.83 for non-randomly sampled knockout strains (DE5), and 0.61-0.9 for *mreB* and *mrdA* single point mutation strains. From these plots, we see unusually low correlation of plate DE3c with its replicates, and we decided to remove it from further analysis. Inspecting the colonies grown on plate DE3c showed that the colonies tended to be larger and less thick than those of the other replicate plates. Thus, we hypothesize that its discrepancy from its replicates may be due to higher plate moisture. Figure 2.S3b shows that the Pearson correlation coefficient for strains grown in different positions on the same plate was 0.55-0.67. Thus, we observe comparable levels of variation due to differences in a strain's position on the plate and differences between plates.

We also tested for variation that may result due to the noisiness of averaging a finite number of bead trajectories for each colony. For the selected knockout strains, we randomly split the bead trajectories in each colony's field of view in half, and calculated the MSD separately for each set of trajectories. Figure 2.S3c shows that the Pearson correlation coefficient between the two random sets of half of the trajectories is 0.78. Thus, we see that while finite track numbers can lead to variation between replicates, it does not play the largest role in determining variation between replicates, as differences between plates generally leads to more variation.

As described in the main text, in order to correct for variations across plates when measuring the distribution of demographic noise effects, we grew 8 wild type colonies on each plate with the randomly selected knockout strains, varying their positions on each plate. We normalized the measurements for each colony to the average wild type value on that plate for all measurements (bead MSD and phenotypic traits) to get a relative measurement. This effectively removes any systematic differences between plates but does not remove additional non-systematic variation from differences between plates, variation between positions, or variation from the number of tracks in each field of view.

Beneficial sectors in monoculture colonies

We filtered out colonies with a noticeable beneficial sector in the field of view of the timelapse from the analysis as these beneficial mutants may change the measured strength of demographic noise if grown to a large enough size. We find no bias in the mean squared displacement measurement for colonies that had beneficial sectors compared to those that did not (Figure 2.S4). We identify beneficial sectors first by looking for bead trajectories that are strongly diverging in space, indicating faster growth than surrounding cells. Next, we verify this in the brightfield timelapse by looking for bulges at the colony front that expand over time. The majority of the 161 genotypes that are filtered out of the 352 randomly selected knockouts are due to two or more replicates having beneficial sectors.

To better understand if the number of colonies with beneficial sectors is surprising, we can estimate the expected number of de novo beneficial mutations from growth in the colony:

$$m_{b,est} = \mu_b p_{est} N \left(\frac{T}{\tau_{gen}} \right) \tag{2.S4}$$

where μ_b is the beneficial mutation rate per genome per generation, p_{est} is the establishment probability, N is the number of cells (for the colony, it's the number of cells in the front layer that are actively dividing), T is the total time, and τ_{gen} is the generation time. Using $\mu_b = 10^{-5}$, $p_{est} > 10^{-3}$ (Ref [26]), $N = 10^9$, $T/\tau_{gen} = 10$, then the expected number of established de novo beneficial mutations from growth in a single colony is at least $m_{b,est} > 100$, which means that we should not be surprised to see de novo beneficial mutations that have established in the colonies. We also note that the number of colonies identified with beneficial sectors from the imaging data is an underestimate of the true number of beneficial mutations in the entire colony because we are not able to identify weak beneficial mutations or mutations that arise later on and stay a small size and we also only image about 1/10 of the colony front.

An additional contribution to the number of beneficial sectors could be standing variation in the glycerol stock. This was tested by streaking single colonies from the glycerol stock, and picking colonies to inoculate the culture as a condition without standing variation. Figure 2.S5 shows that cultures started from single colonies exhibited beneficial sectors as well as those started from glycerol stock. Thus the beneficial sectors that we see are likely a combination of both de novo mutations and standing variation from glycerol stock.

Gene enrichment analysis

We asked whether there are gene categories that are associated with extreme changes to demographic noise. To do this, we compared the weighted median MSD of gene knockouts associated with a particular GO or KEGG term with randomly chosen MSDs from the knockout distribution.

From this analysis, we did not find any significant GO terms, and we only found one significant KEGG term that was enriched for higher MSD than the weighted median of the

entire knockout distribution: ATP-binding cassette transporters. It is possible that expanding the dataset by measuring more genotypes or decreasing the amount of technical noise by performing more experimental replicates may result in more hits in the gene enrichment analysis. However, another possibility for the few gene enrichment hits is that the influence of phenotypic traits on demographic noise dominates over any differences between gene categories.

Additional measurements of distribution of drift effects

In the main text we presented results for the distribution of demographic noise for a randomly selected set of 191 single gene knockout strains, in order to get a representative of the distribution for all single gene knockouts. To test for extreme differences in demographic noise, we also enriched for strains that we hypothesized to significantly change demographic noise. This included (1) selected single gene knockout strains that had altered biofilm forming ability in liquid culture [50] or different cell shapes [1], and different *E. coli* strain backgrounds, and (2) *mreB* and (3) *mrdA* single point mutant libraries that were enriched for cell shape differences [57].

Figure 2.S18 shows the distributions of demographic noise of the selected single gene knockouts, the *mreB* point mutants, and the *mrdA* point mutants, compared with the wild type BW25113 and the randomly selected knockouts. Note that the *mreB* and *mrdA* strains are on the MG1655 strain background, but here we will compare them to the wild type BW25113 measurements because we only made 2 measurements of MG1655 which is not enough to compare distributions and we found that MG1655 has a similar strength of demographic noise to BW25113 (Figure 2.1e).

The tail at the lower end of the distributions for all three additionally tested sets of strains does not overlap with the gaussian fit to the wild type distribution nor do they completely overlap the grayed out region from the finite number of wild type measurements. The selected single gene knockout distribution and the *mreB* point mutants distribution are significantly different from that of the wild type using a two-sample Kolmogorov-Smirnov test ($p = 1.9 \times 10^{-10}$ for selected knockouts, 6×10^{-5} for *mreB* point mutants), while the *mrdA* point mutant library is less significantly different from the wild type distribution ($p = 6.8 \times 10^{-2}$), most likely due to the fewer number of measurements.

The distributions of the strength of demographic noise for the selected single gene knockouts and the *mreB* point mutants are wider than that of the randomly selected knockouts. Some strains even have MSD close to zero. Thus, we see that the maximally allowed changes to demographic noise are much more extreme than we have sampled with our randomly selected strains. Sampling more strains will also lead to a more extreme knockout distribution. Similarly to what's observed for the randomly selected knockouts, many more mutants have a lower strength of demographic noise than a higher strength of demographic noise compared to the wild type.

Comparison of cell shape between growth in liquid culture and as a colony

We used existing datasets for the cell shape of single gene knockout strains in the Keio collection [61, 17]. However, cells in these previous studies were grown in liquid culture, and it is unclear whether the cell shape could be different in colonies (the condition in this study). To determine whether these datasets would be applied to our experiments, we measured cell shape from growth in liquid culture and colonies for a subset of 12 strains (see Supplementary methods).

Figure 2.S13c shows a Pearson correlation coefficient between the mean circularity of single cells between the colony and liquid culture growth conditions to be 0.46. However, this is primarily driven by the large circularity of the genotype $\Delta gpmI$. Removing it from the measurement of the correlation coefficient gives a low correlation of -0.1 for the remaining genotypes. Thus, in interpreting the lack of correlation between cell shape and bead MSD in the single gene knockouts, we cannot rule out that it is because the cell shape is different in the colonies than the liquid culture condition of the previous datasets. We also note that the two previous datasets [61, 17] in liquid culture do not have any correlation with one another, further suggesting cell shape is sensitive to the precise growth conditions (Figure 2.S13d). We note that the dataset from Ref. [61] did not reproduce the extreme cell shapes in the current images on the Keio collection database, which may be because the images have been changed online since their analysis.

Non-neutral experiments

Figure 2.S9a shows the fitness difference of the resistant and susceptible pairs in different strain backgrounds as measured by a colony collision assay (see Methods). Interestingly, different strain backgrounds exhibit different fitness coefficients despite having the same plasmids. This may be due to differences in plasmid copy number or epistatic effects between the plasmid and the rest of the genome. Studying the interaction between the strain background and fitness due to an antibiotic resistance gene on a plasmid would be an interesting avenue for future work.

We tested multiple different initial mutant fractions in order to be able to resolve individual sectors for different strain backgrounds and fitnesses. Figure 2.S17b shows the establishment probabilities for three different initial fractions of the resistant type that are approximately $p_i = 0.002$, 0.005, and 0.02. Figure 2.S9b shows the actually measured initial fractions for each of these genotypes using colony counting; the general trend of initial fractions matches the expectation, but there are offsets between strains. While the ordering of the establishment probabilities of the strains is similar across initial mutant fractions, the values themselves are slightly different, with slightly higher establishment probabilities for a lower initial mutant fractions. For better statistics, we used data from all initial mutant fractions to generate Figure 2.4b. The largest contribution to experimental error is in the measurement of the initial mutant fraction due to poisson sampling from counting a finite

number of CFUs. Additional noise in the data may result from slightly different chloramphenicol concentrations between the agar plates used for the fitness measurements and the plates used for establishment probability measurements.

Supplementary figures

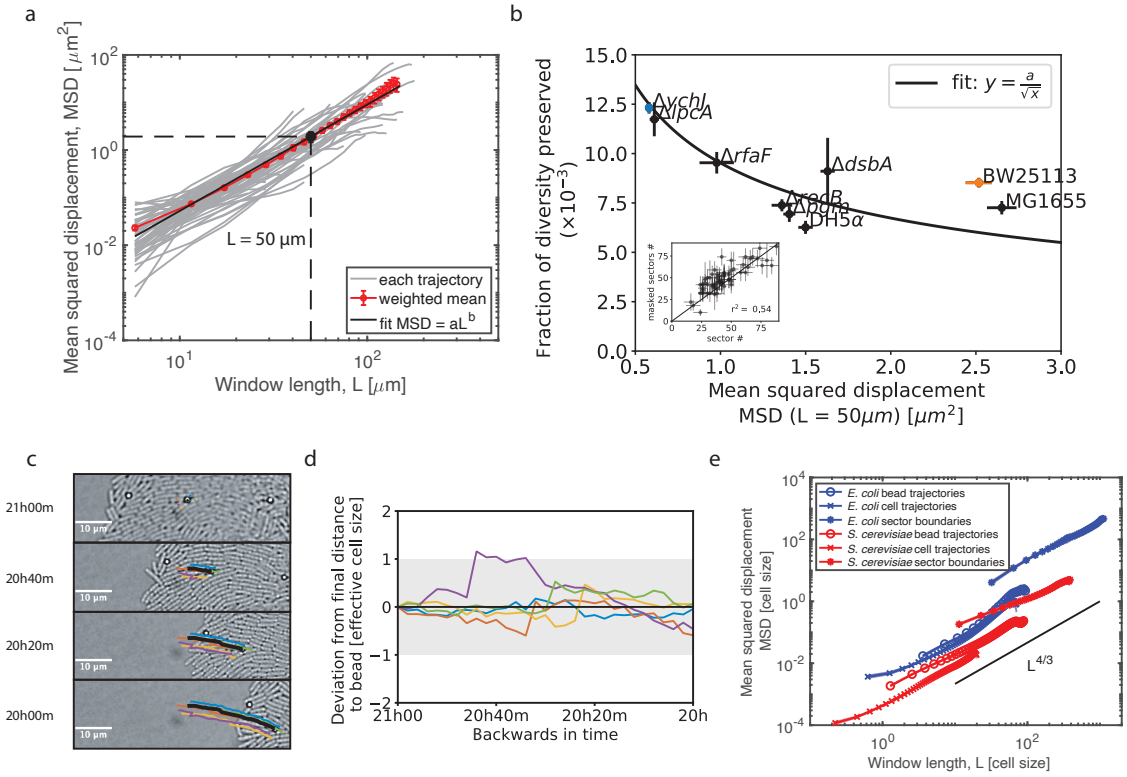

Figure 2.S1: Bead-based sparse lineage tracing method. (a) Example of measurement of mean squared displacement (MSD) for a single colony (a single field of view). Gray lines show the MSD for single bead trajectories. The red line shows the average across all trajectories weighted by the inverse squared error of MSD for each trajectory at each window size. Error bars represent the standard error in the weighted mean. The black line shows the weighted least squares fit of the mean to a power law. The summary statistic MSD(L = 50 μm) is calculated by interpolating the fit to L = 50 μm. (b) MSD(L = 50 μm) compared to the fraction of diversity preserved when colonies are masked by the outline of the smallest colony to account for growth rate differences between colonies. Inset shows a comparison of the number of sectors counted with and without masking. (c) The trajectory of a single bead (black) behind the front and the lineages of cells neighboring it in the final timepoint (colors) traced backwards in time over 1 hour in wild type strain BW25113. (d) The deviation of the distance between the cell lineages and the bead from the final distance, backwards in time. Colors are the same as in (c). All cells neighboring the bead in the latest timepoint are neighboring the bead in the earliest timepoint, except for the yellow lineage (even though the yellow lineage does stay within a single cell width of its final distance to the bead). (e) *E. coli* has higher MSD than *S. cerevisiae* in all three methods for measuring demographic noise. Sector boundary mean squared displacement was measured according to [28]

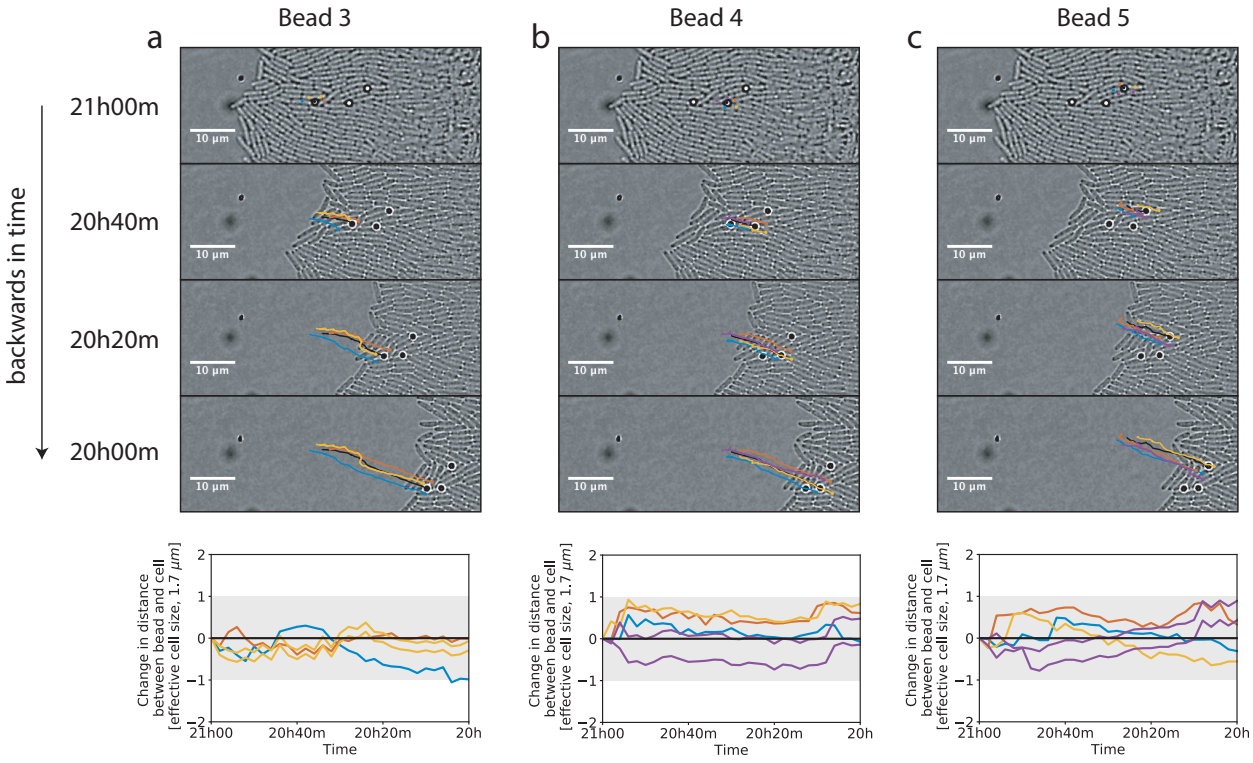

Figure 2.S2: The trajectory of three single beads (black) and the lineages of the cells neighboring it in the final-timepoint (colors) traced backwards in time in the Keio collection wild type strain. The panels below correspond to the trajectories from the time series images directly above and show the deviation of the distance between the cell lineages and the bead from the final distance, backwards in time. Colors are the same as in the time series images. The gray shaded region shows a single cell width away or towards the bead. All cells that neighbor the bead in the final timepoint are neighbors of the bead in the first timepoint (with the exception of the purple lineage in c) and stay within a single cell width of the final distance to the bead.

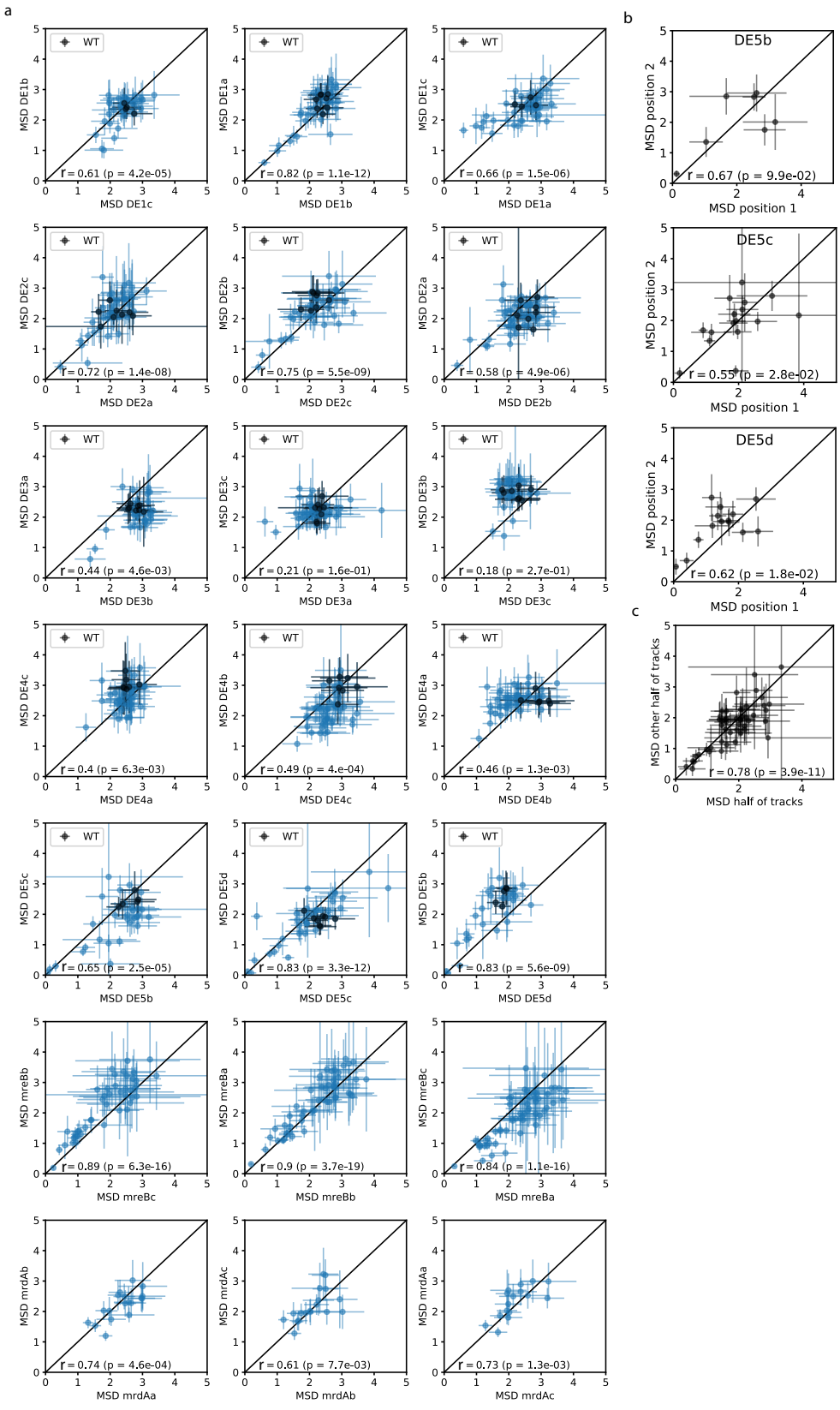
a
WT
MSD DE1b
MSD DE1c
r = 0.61 (p = 4.2e-05)
MSD DE1a
MSD DE1b
r = 0.82 (p = 1.1e-12)
MSD DE1c
MSD DE1a
r = 0.66 (p = 1.5e-06)
MSD DE2c
MSD DE2a
r = 0.72 (p = 1.4e-08)
MSD DE2b
MSD DE2c
r = 0.75 (p = 5.5e-09)
MSD DE2a
MSD DE2b
r = 0.58 (p = 4.9e-06)
MSD DE3a
MSD DE3b
r = 0.44 (p = 4.6e-03)
MSD DE3c
MSD DE3a
r = 0.21 (p = 1.6e-01)
MSD DE3b
MSD DE3c
r = 0.18 (p = 2.7e-01)
MSD DE4c
MSD DE4a
r = 0.4 (p = 6.3e-03)
MSD DE4b
MSD DE4c
r = 0.49 (p = 4.e-04)
MSD DE4a
MSD DE4b
r = 0.46 (p = 1.3e-03)
MSD DE5c
MSD DE5b
r = 0.65 (p = 2.5e-05)
MSD DE5d
MSD DE5c
r = 0.83 (p = 3.3e-12)
MSD DE5b
MSD DE5d
r = 0.83 (p = 5.6e-09)
MSD mreBb
MSD mreBc
r = 0.89 (p = 6.3e-16)
MSD mreBa
MSD mreBb
r = 0.9 (p = 3.7e-19)
MSD mreBc
MSD mreBa
r = 0.84 (p = 1.1e-16)
MSD mrdAb
MSD mrdAa
r = 0.74 (p = 4.6e-04)
MSD mrdAc
MSD mrdAb
r = 0.61 (p = 7.7e-03)
MSD mrdAa
MSD mrdAc
r = 0.73 (p = 1.3e-03)
b
DE5b
MSD position 2
MSD position 1
r = 0.67 (p = 9.9e-02)
DE5c
r = 0.55 (p = 2.8e-02)
DE5d
r = 0.62 (p = 1.8e-02)
c
MSD other half of tracks
MSD half of tracks
r = 0.78 (p = 3.9e-11)

Figure 2.S3 *(preceding page)*: Sources of variation (a) Comparison of MSD for strains grown in the same position on different plates (r = 0.4-0.82 for randomly sampled knockout strains DE1-4, r = 0.65-0.83 for non-randomly sampled knockout strains DE5, and r = 0.61-0.9 for *mreB* and *mrdA* single point mutants). Wild type strains on each plate indicated in black. The plate DE3c exhibited low correlation with its replicates and was removed from further analyses. (b) Comparison of MSD for strains grown in different positions on the same plate (r = 0.55-0.67). (c) Comparison of MSD from randomly splitting the bead trajectories from a single colony in half and separately calculating MSD for each set of trajectories (r = 0.78).

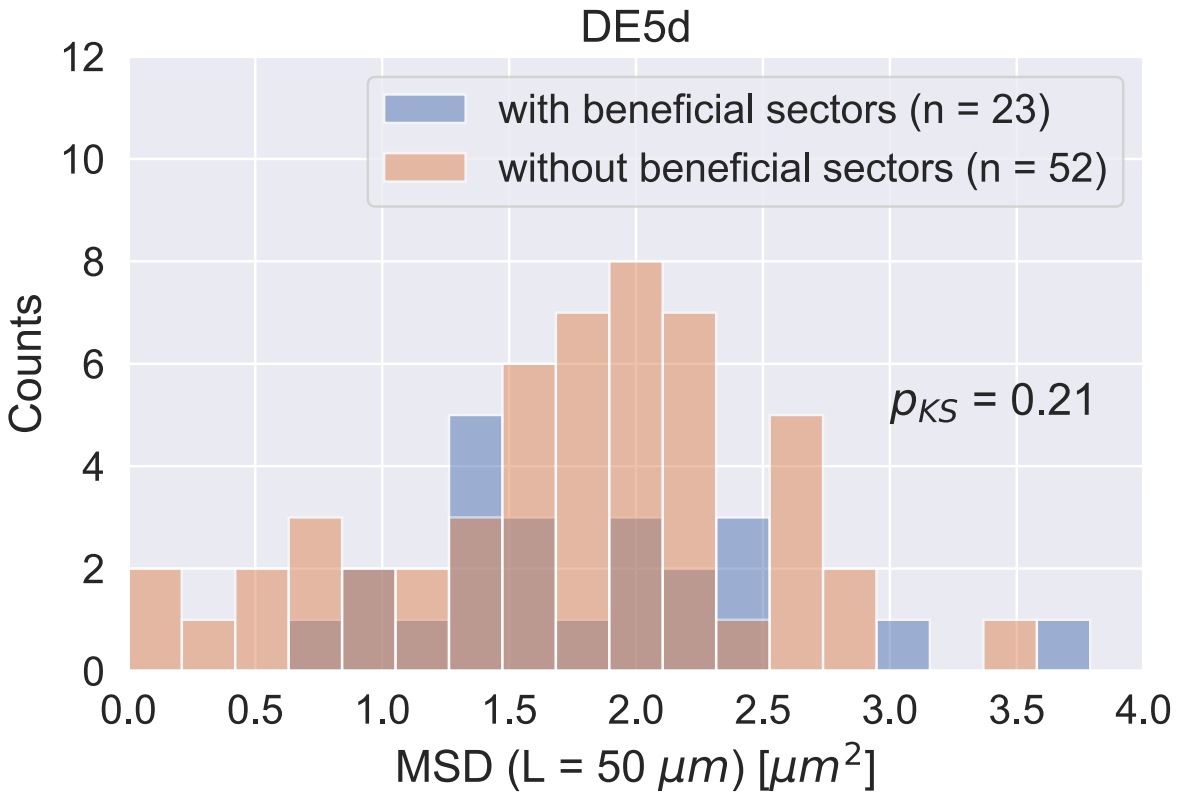

Figure 2.S4: The distribution of mean squared displacements of colonies with and without beneficial sectors. Using a two-sided Kolmogorov-Smirnov test, the distributions are not found to be significantly different (p = 0.21).

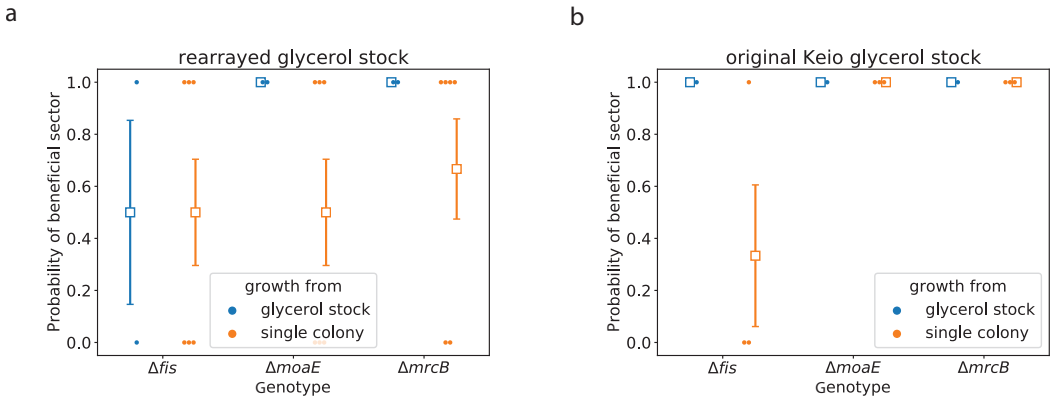

Figure 2.S5: Testing for standing variation vs de novo mutations. Probability of seeing at least 1 beneficial sector in colonies grown from (a) the rearrayed glycerol stock or (b) the original Keio collection glycerol stock received from the National BioResource Project. Colonies were grown from liquid culture that was inoculated either directly by scraping off some glycerol stock (blue) or by first streaking the glycerol stock onto a plate, and picking from a single colony (orange). Otherwise, the experimental conditions were the same as that for measuring the distribution of demographic noise. Dots show individual colonies with none (0) or at least 1 beneficial sector (1) and squares show the average across colonies, which represents the probability of seeing a beneficial sector in a given condition. The error bars represent the standard deviation from binomial sampling. In all conditions, beneficial sectors were observed in at least one colony. This supports the hypothesis that beneficial mutations can arise de novo on the timescale of the experiment. Most likely, both beneficial mutations and standing variation from the glycerol stock contribute to the observation of beneficial sectors in the main experiment.

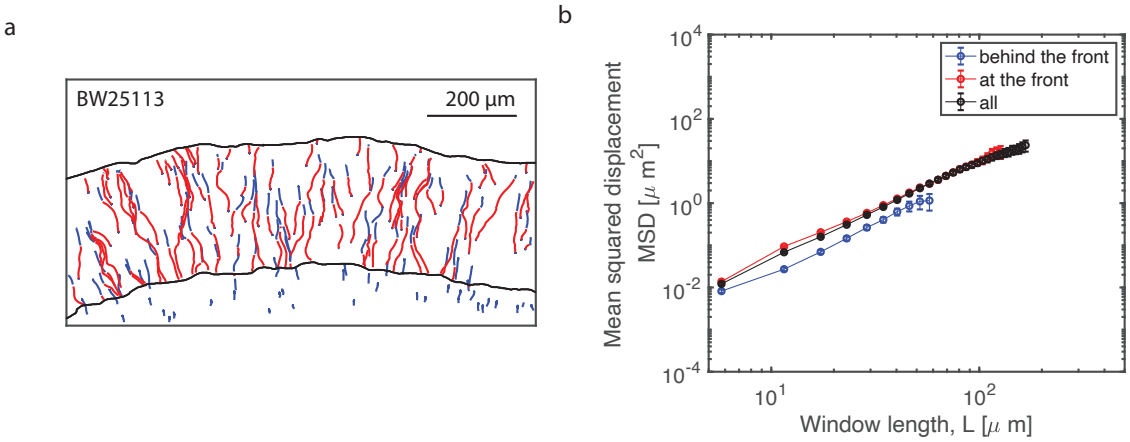

Figure 2.S6: MSD of beads at and behind the front. (a) Trajectories of beads at the front (red, measured as within 6μm of the front) and behind the front (blue). (b) Beads at the front give similar MSD to that of overall bead trajectories.

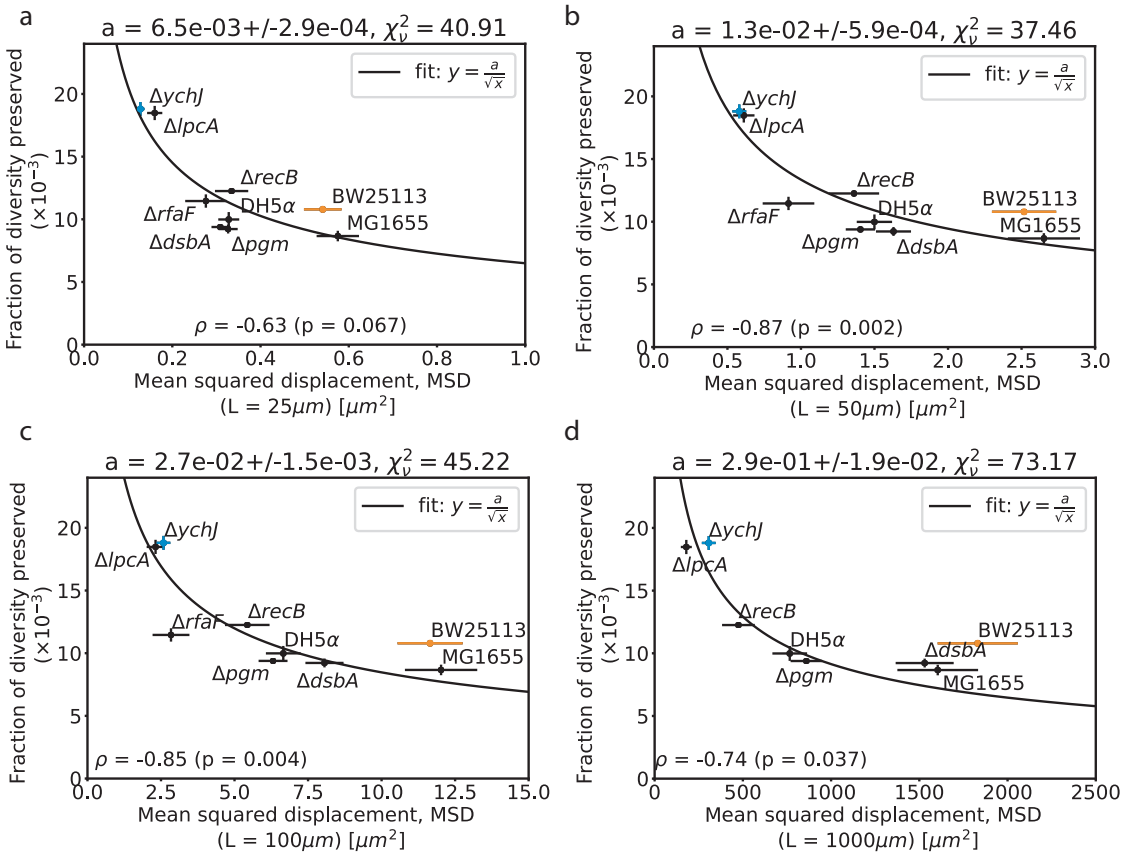

Figure 2.S7: Selection of window size for the summary MSD statistic. Comparison of the fitted MSD value at different window sizes to the fraction of diversity preserved in a colony grown from a neutral mixture of two fluorescent strains for 24 hours. Error bars in MSD represent the standard error of the weighted mean (7-8 colonies) where weights come from uncertainties in the fit of MSD as a function of L to a power law (see Methods) and error bars in the fraction of diversity preserved represent the standard error of the weighted mean (8 colonies) where weights come from uncertainties in counting the number of sectors . Fit to $y = ax^{-1/2}$, where x is the MSD and y is the fraction of diversity preserved, shows that the window size of $L = 50$ μm gives the lowest chi-squared value.

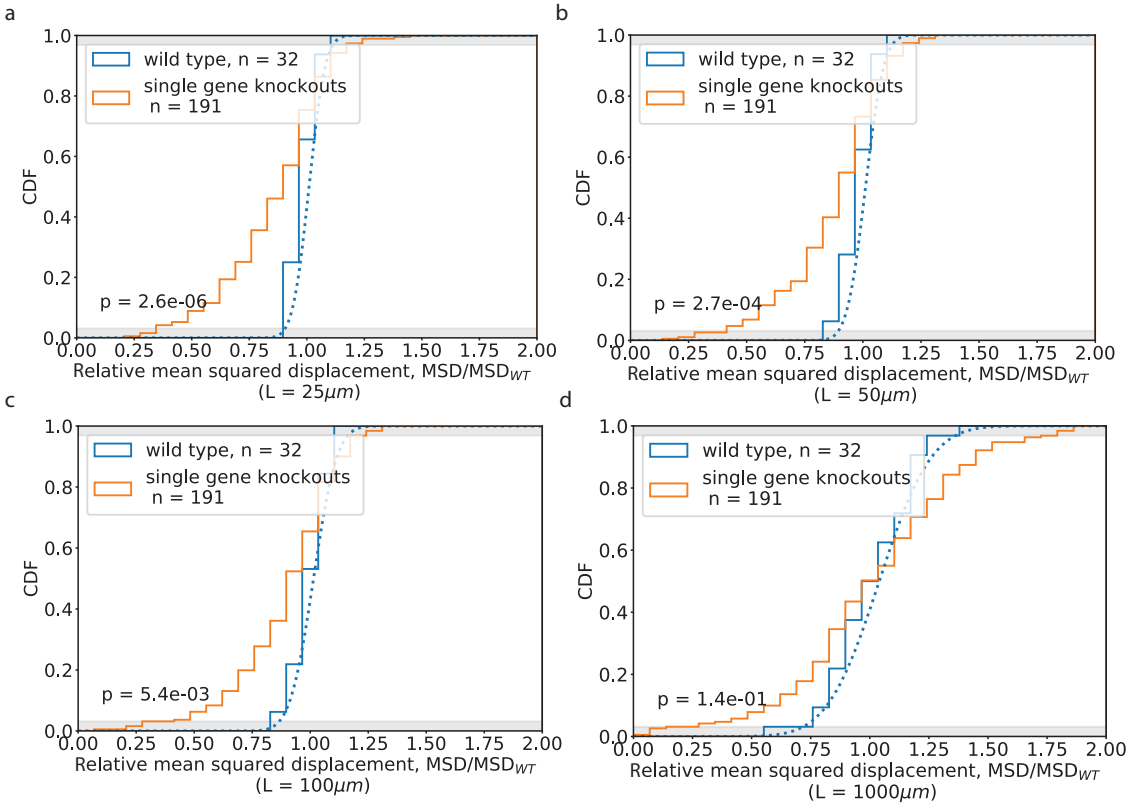

Figure 2.S8: Distribution of demographic noise effects for the MSD reported at different window sizes. The blue dotted line shows a gaussian fit to the wild type distribution. The tails of the gaussian fit do not overlap with the tails of the knockout distribution. The gray shaded region shows [1/number of wild type measurements] and [1-(1/number of wild type measurements)], which is the limit of the resolution of the wild type distribution being compared to. p values show the probability that the wild type and knockout distributions are the same using a two sample Kolmogorov-Smirnov test. The knockout distribution is different from the wild type distribution to $p < 0.05$ for all window sizes except $L = 1000\mu$m.

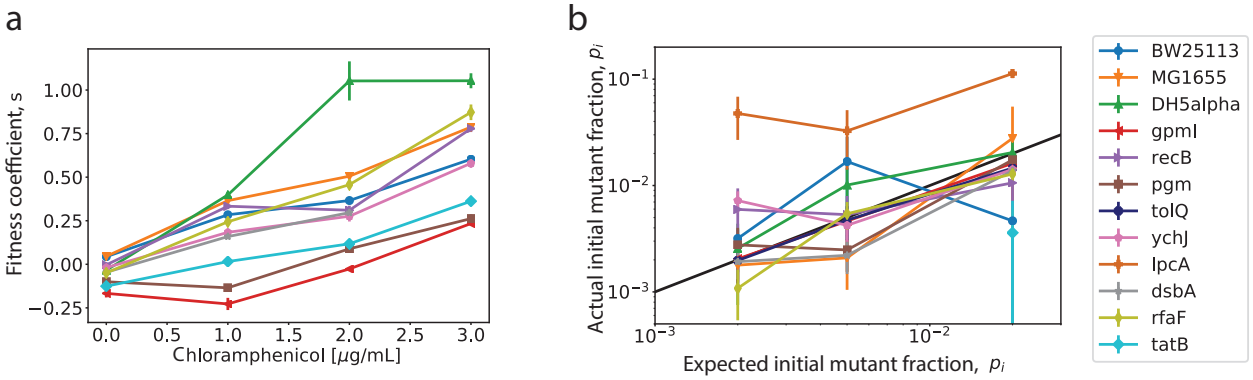

Figure 2.S9: (a) The fitness coefficient (s) of a chloramphenicol resistant strain compared to chloramphenicol sensitive strain across strain backgrounds as measured on plates using a colony collision assay (Methods). (b) Expected and actual initial fraction of the resistant strain. Black line shows 1:1 relationship. $\Delta tatB$ had a large error in the measurement of the initial fraction of the mutant type from counting CFUs and was removed from Figure 2.4 in the main text and Figure 2.S17.

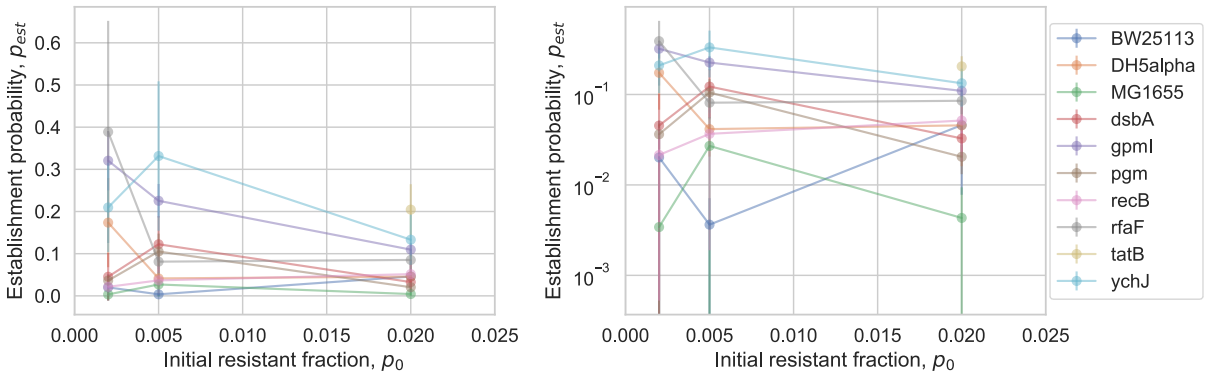

Figure 2.S10: The establishment probability as a function of the initial resistant (mutant) fraction on a linear scale (left) and log scale (right) for an interpolated fitness coefficient of s = 0.15. There is no clear downward bias in the establishment probability as a function of the initial resistant fraction, suggesting that sector coalescences are unlikely for these experimental parameters.

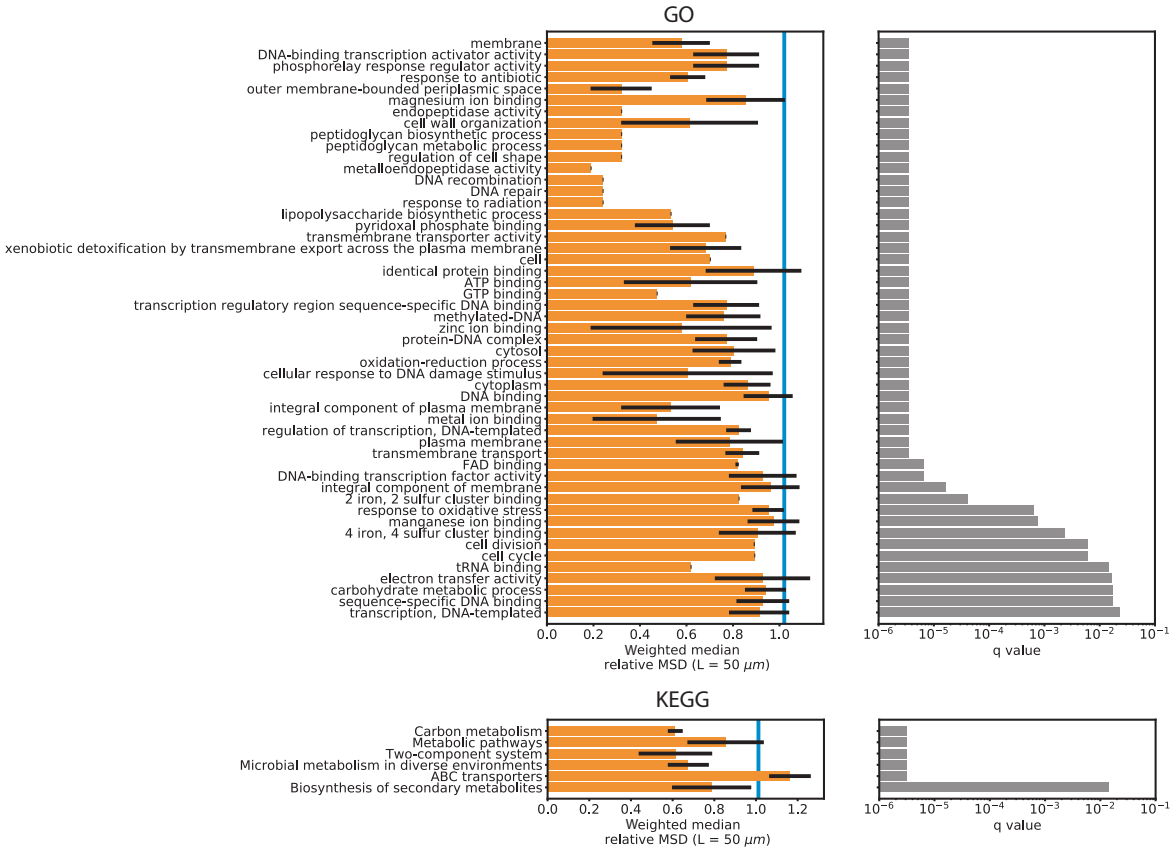

Figure 2.S11: GO and KEGG terms with significantly different MSD from that of the WT at 5% level. The significance of each term was tested by randomly drawing the same number of measurements from the wild type distribution as the number of genotypes associated with that term in our KO subset. We repeated this 106 times and calculated the fraction of times that the weighted median of the random WT subset was more extreme than that of the true values associated with that term. The p value was calculated as twice this fraction (because in determining whether a value is more extreme, we first check whether the weighted median MSD of the term is greater than or less than the weighted median MSD of the full WT distribution). (Left) Weighted median MSD of significant terms and WT weighted median MSD (blue line). Error bars represent the mean absolute deviation. (Right) q-values after applying Benjamini-Hochberg FDR correction to account for multiple testing.

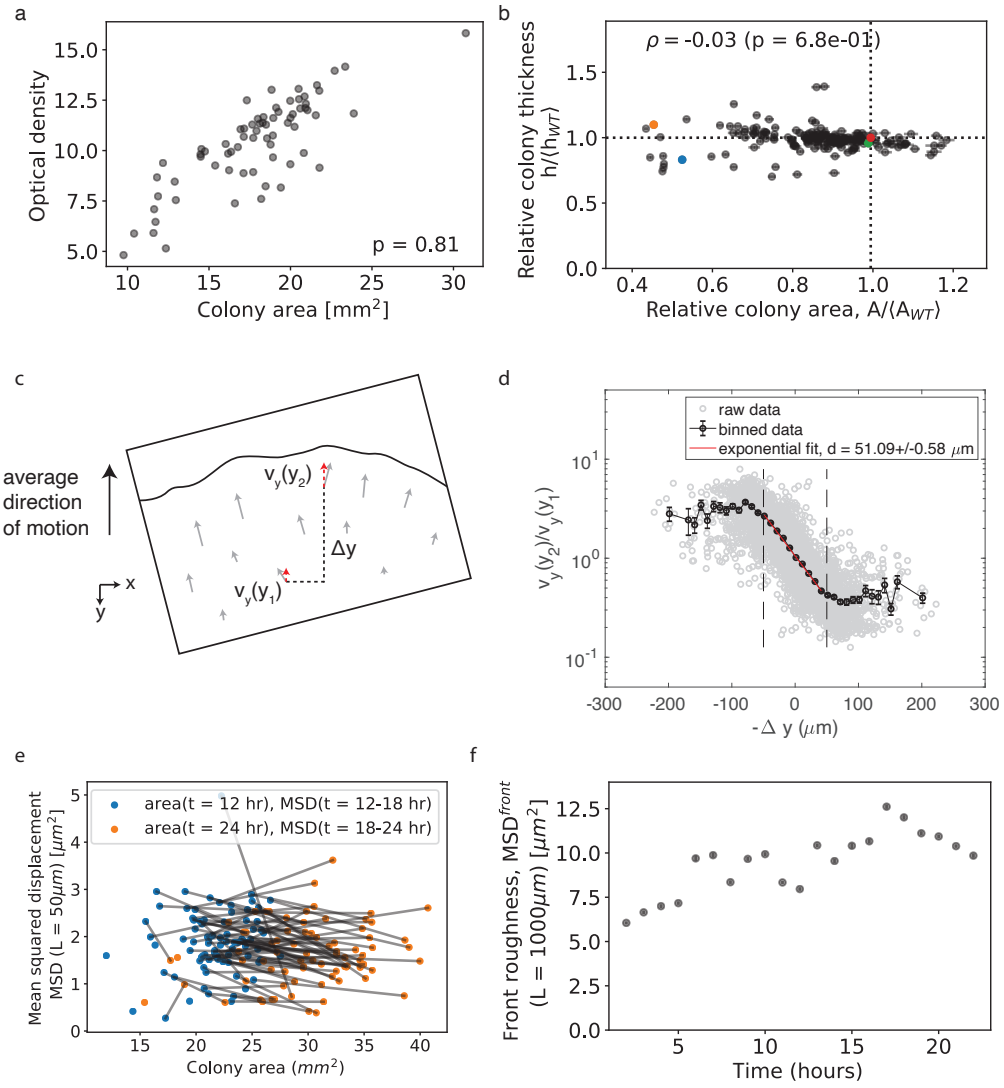
a
Optical density
Colony area [mm^2]
p = 0.81
b
ρ = -0.03 (p = 6.8e-01)
Relative colony thickness h/⟨h_{WT}⟩
Relative colony area, A/⟨A_{WT}⟩
c
average direction of motion
$v_y(y_2)$
$v_y(y_1)$
Δy
x
y
d
raw data
binned data
exponential fit, d = 51.09+/-0.58 μm
$v_y(y_2)/v_y(y_1)$
-Δ y (μm)
e
area(t = 12 hr), MSD(t = 12-18 hr)
area(t = 24 hr), MSD(t = 18-24 hr)
Mean squared displacement MSD (L = 50μm) [μm^2]
Colony area (mm^2)
f
Front roughness, MSD^{front} (L = 1000μm) [μm^2]
Time (hours)

Figure 2.S12 *(preceding page)*: Additional tests of phenotypic traits (a) Colony area is a highly correlated measure of the optical density of the colony resuspended in liquid, which gives the total (alive and dead) biomass. (b) Colony thickness (see Methods) is not correlated with colony area, showing that area is a good measure of growth, rather than just spreading on the surface of the plate. (c) Schematic of measurement growth layer depth using bead displacements from a single colony's field of view. For a consecutive pair of timepoints, Δy gives the distance between a pair of beads along the average direction of motion, and and $v_y(y_1)$ and $v_y(y_2)$ give the bead displacements projected onto the average direction of motion. (d) Gray points show measurements for all pairs of beads across all pairs of consecutive time points. Black points give binned measurements. Error bars represent standard error of the mean. The binned points roughly followed an exponential decay function between Δy = -50μm and Δy = 50μm, and we fit to an exponential decay function within this range, as measurements of bead pairs that are farther away may be dominated by noise. The decay length, d, is extracted from the fit and taken as the growth layer depth. (e) The MSD for bead trajectories from 12-18 hours as a function of the area of the colony at 12 hours (blue points) and the MSD for bead trajectories from 18-24 hours as a function of the area of the colony at 24 hours (orange points). The same colony over time is connected by a gray line. Bead trajectory MSD mostly decreases with increasing colony area over time. Thus, the positive correlation between colony area and MSD seen across genotypes cannot be explained by changes in the bead trajectories over time. (f) Colony front roughness (see Methods) for *E. coli* DH5α over time. Colony front roughness increases initially but levels off around 19 hours. Since we measure the colony front roughness at 24 hours, we expect to have passed the time when the behavior of the front roughness is transient. Note that this measurement was done on a colony grown in a 10cm diameter petri dish rather than an omniplate, and the colony may have access to a different total amount of nutrients. However, while the growth condition does not exactly match that in the experiment, we used this data to better understand the approximate timescales of when front roughness saturates in time.

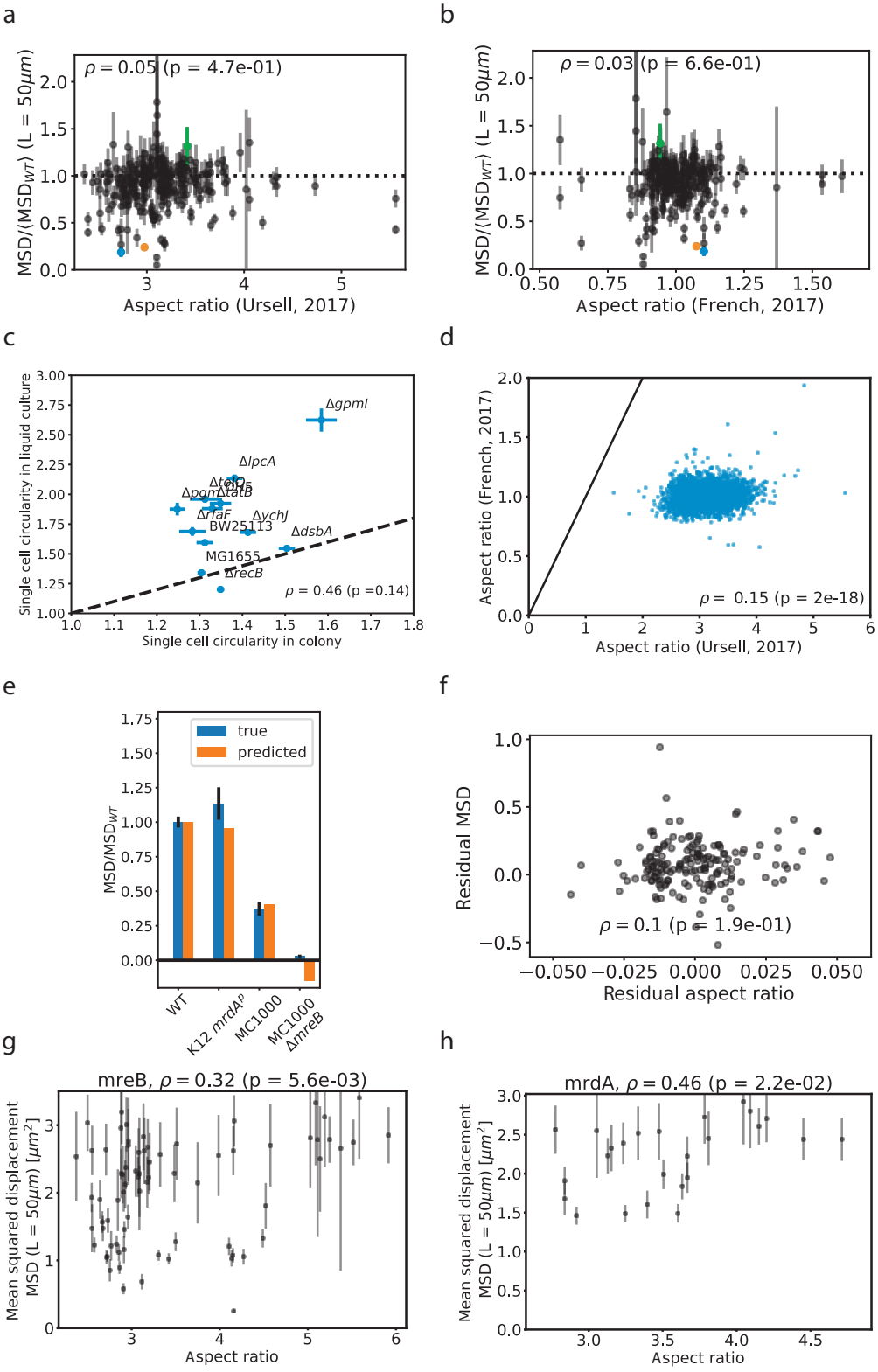
a
ρ = 0.05 (p = 4.7e-01)
MSD/⟨MSD_WT⟩ (L = 50μm)
Aspect ratio (Ursell, 2017)
b
ρ = 0.03 (p = 6.6e-01)
MSD/⟨MSD_WT⟩ (L = 50μm)
Aspect ratio (French, 2017)
c
Single cell circularity in liquid culture
Single cell circularity in colony
Δgpml
ΔlpcA
ΔpgmΔtatB
ΔrfaF
ΔychJ
BW25113
ΔdsbA
MG1655
ΔrecB
ρ = 0.46 (p =0.14)
d
Aspect ratio (French, 2017)
Aspect ratio (Ursell, 2017)
ρ = 0.15 (p = 2e-18)
e
true
predicted
MSD/MSD_WT
WT
K12 mrdA^p
MC1000
MC1000 ΔmreB
f
Residual MSD
Residual aspect ratio
ρ = 0.1 (p = 1.9e-01)
g
mreB, ρ = 0.32 (p = 5.6e-03)
Mean squared displacement MSD (L = 50μm) [μm²]
Aspect ratio
h
mrdA, ρ = 0.46 (p = 2.2e-02)
Mean squared displacement MSD (L = 50μm) [μm²]
Aspect ratio

Figure 2.S13 *(preceding page)*: Tests of correlation between cell shape and demographic noise. The strength of demographic noise is uncorrelated with cell aspect ratio in the 191 randomly sampled strains from the Keio collection and 34 selected strains from the Keio collection with cell shape data taken from (a) Ursell et al [61] and (b) French et al [17]. (c) The circularity (see Supplementary methods) of cells grown in liquid culture compared to those grown in a colony. Dashed black line depicts 1:1 relationship. The Pearson correlation coefficient is 0.46 with all genotypes, but only -0.1 when excluding $\Delta gpmI$. The error bars represent the standard error of the mean across the 50-100 cells measured per genotype. (d) Comparing cell shape data from Ursell et al and French et al across all single gene knockouts in the Keio collection. (e) Relative bead trajectory MSD in cell shape mutants and those predicted by the best fit Lasso regression model from the main text which primarily includes colony-level traits. (f) Partial correlation of cell aspect ratio when controlling for all other phenotypes for the random set of single gene knockouts from the Keio collection. Correlation of demographic noise to single cell aspect ratio in $mreB$ (g) and $mrdA$ (h) single point mutants is higher than that seen for the single gene knockouts, possibly because these strains were enriched for cell shape differences. Strains and data from Ref [57].

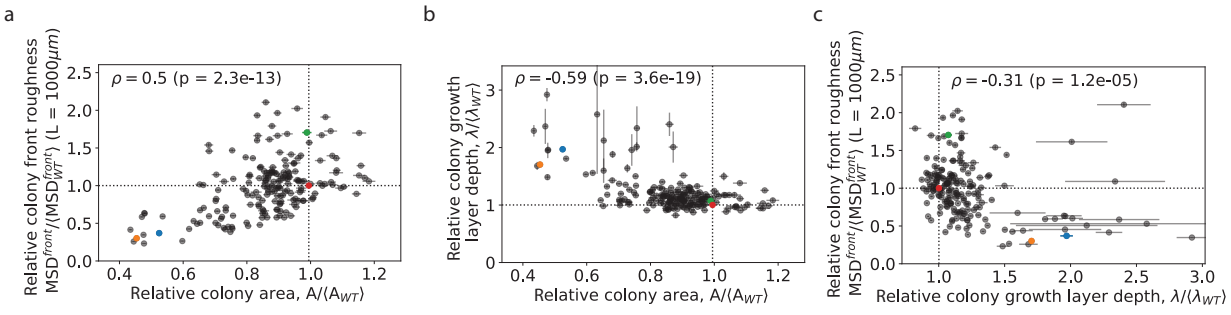

Figure 2.S14: Correlation of colony traits with one another.

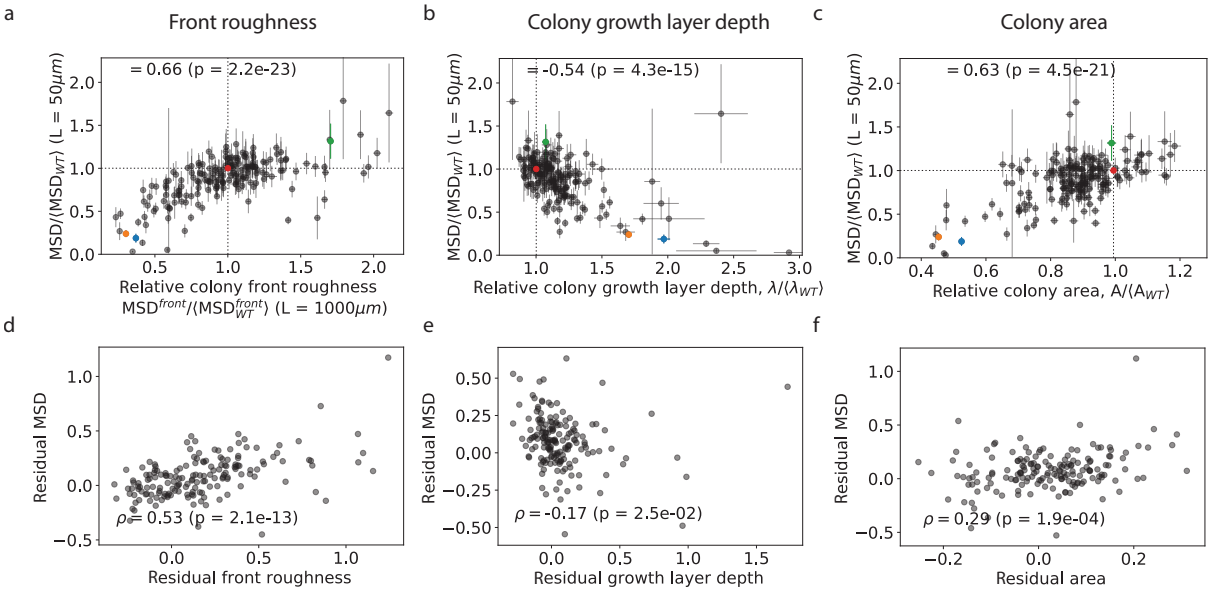

Figure 2.S15: Correlation between the bead MSD and (a) front roughness, (b) growth layer depth, or (c) colony area (same as in Figure 2.3). Below each plot is the partial correlation between the bead MSD and each phenotypic trait (after controlling for correlations with all other traits).

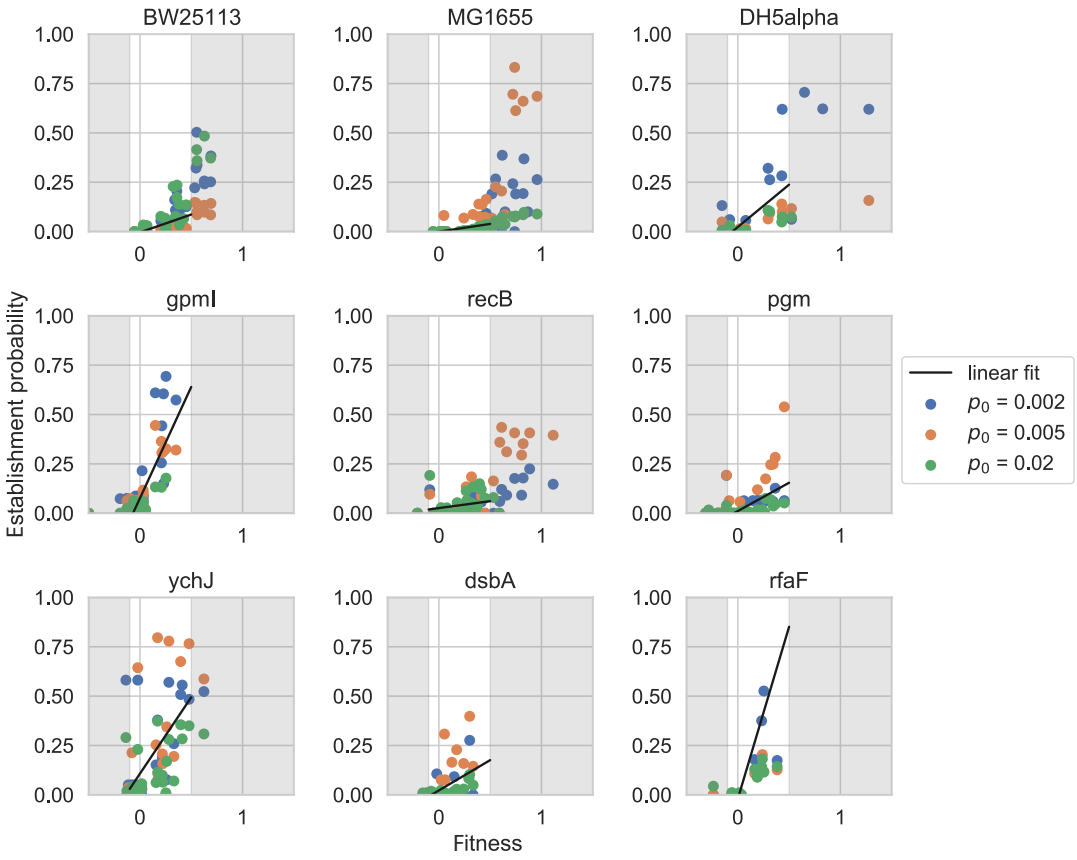

Figure 2.S16: The establishment probability as a function of the fitness coefficient of a chloramphenicol resistant mutant for 9 selected strain backgrounds. Points that fall within fitness coefficients $-0.1 < s < 0.5$, where the colony collision assay is valid, are fit linearly (black line).

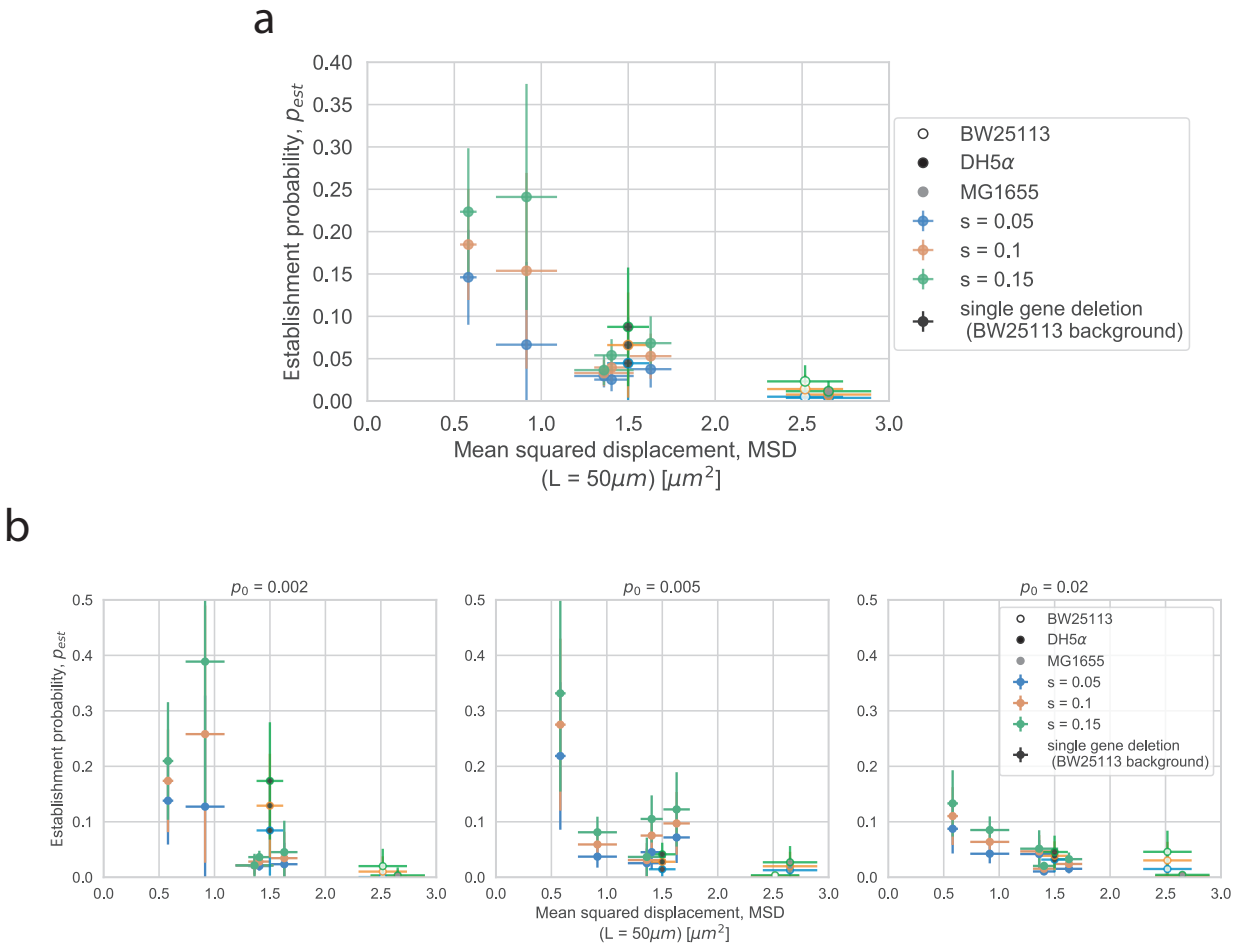

Figure 2.S17: Fitted establishment probability at three different fitness coefficients from fitting (a) all initial resistant fractions together and (b) each initial resistant fraction separately as a function of bead trajectory MSD for 5 selected single gene deletion strains and 3 wild type strains. Error bars in the establishment probability represent linear fitting error (see Supplementary methods) and and error bars in MSD represent the standard error of the weighted mean (N = 7-8, see Methods).

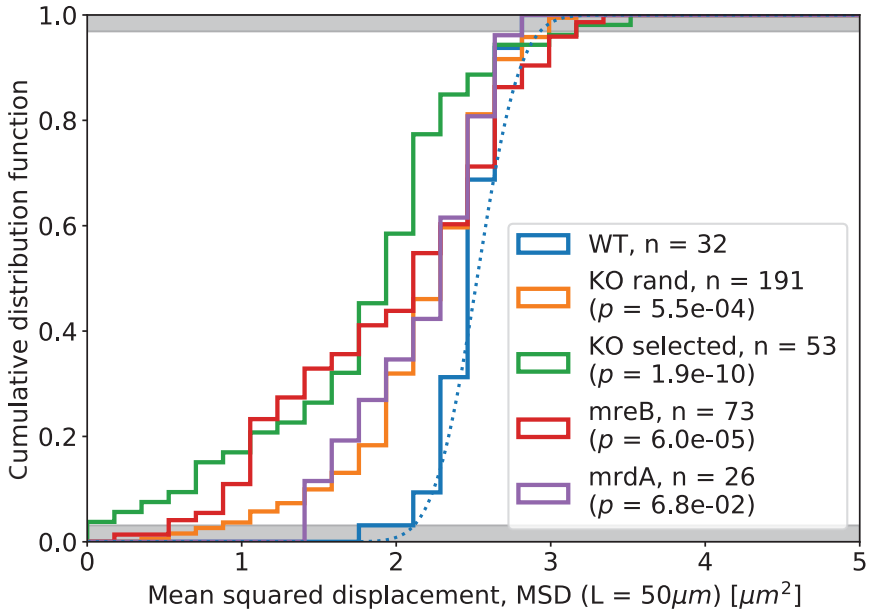

Figure 2.S18: Additional measurements of demographic noise effects in non-randomly selected strains. (a) The distribution of strengths of demographic noise for the wild type BW25113 (WT), randomly selected single gene knockouts (KO rand), specifically selected single gene knockouts and strain backgrounds (KO selected), *mreB* single point mutants (*mreB*), and *mrdA* single point mutants (*mrdA*). All except *mrdA* single gene knockouts show a significantly different distribution of demographic noise effects compared to the WT to a significance level of $p < 0.05$.

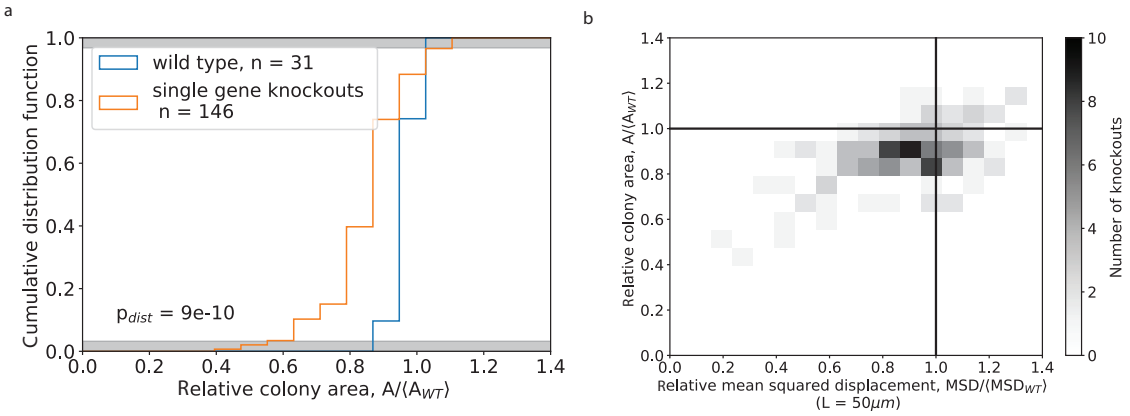

Figure 2.S19: Joint distribution of demographic noise and colony fitness as measured by colony area. (a) The distributions of colony areas for the wild type BW25113 strain and the randomly selected single gene knockout strains. The distributions are significantly different to $p = 9 \times 10^{-10}$ as measured by a two-sample Kolmogorov-Smirnov test. (b) The joint distribution of demographic noise (as measured by bead MSD) and fitness (as measured by colony area) effects. The strength of demographic noise and fitness are correlated suggesting a tradeoff between fitness and demographic noise, with the median of both slightly below that of the wild type (black lines).

Additional supplementary files

Additional supplementary files can be found online at https://www.nature.com/articles/s41396-021-00951-9.

Supplementary movie 1: Movie of fluorescent beads being pushed by growing colony of *E. coli* DH5α.
Type of file: movie

Supplementary table 1: List of *E. coli* single gene deletion strains used from Keio collection and plate positions of all strains tested
Type of file: table

Supplementary table 2: List of *mreB* and *mrdA* single point mutant strains used
Type of file: table

2.8 Bibliography

[1] National BioResource Project Keio collection. https://shigen.nig.ac.jp/ecoli/strain/resource/keioCollection/list.

[2] YPD media. doi:10.1101/pdb.rec12315, 2010.

[3] H. K. Alexander and R. C. MacLean. Stochastic bacterial population dynamics restrict the establishment of antibiotic resistance from single cells. *Proceedings of the National Academy of Sciences*, 117(32):19455–19464, 2020.

[4] Y. Azimzade, M. Sasar, and I. Maleki. Invasion front dynamics in disordered environments. *Scientific Reports*, 10(1):1–10, 2020.

[5] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2(1):2006.0008, Jan. 2006. ISSN 1744-4292. doi: 10.1038/msb4100050. URL https://www.embopress.org/doi/full/10.1038/msb4100050.

[6] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 1995. ISSN 2517-6161. doi: 10.1111/j.2517-6161.1995.tb02031.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x.

[7] G. Birzu, O. Hallatschek, and K. S. Korolev. Fluctuations uncover a distinct class of traveling waves. *Proceedings of the National Academy of Sciences*, 115(16):E3645–E3654, 2018.

[8] G. Birzu, S. Matin, O. Hallatschek, and K. S. Korolev. Genetic drift in range expansions is very sensitive to density dependence in dispersal and growth. *Ecology Letters*, 22(11):1817–1827, 2019.

[9] L. Bosshard, I. Dupanloup, O. Tenaillon, R. Bruggmann, M. Ackermann, S. Peischl, and L. Excoffier. Accumulation of Deleterious Mutations During Bacterial Range Expansions. *Genetics*, 207(2):669–684, Oct. 2017. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.117.300144. URL https://www.genetics.org/content/207/2/669.

[10] L. Chao. Fitness of RNA virus decreased by Muller's ratchet. *Nature*, 348(6300):454–455, 1990.

[11] M. M. Desai, D. S. Fisher, and A. W. Murray. The speed of evolution and maintenance of variation in asexual populations. *Current Biology*, 17(5):385–394, 2007.

[12] M.-C. Duvernoy, T. Mora, M. Ardré, V. Croquette, D. Bensimon, C. Quilliet, J.-M. Ghigo, M. Balland, C. Beloin, S. Lecuyer, et al. Asymmetric adhesion of rod-shaped bacteria controls microcolony morphogenesis. *Nature Communications*, 9(1):1–10, 2018.

[13] A. Eriksson, B. Mehlig, M. Rafajlovic, and S. Sagitov. The Total Branch Length of Sample Genealogies in Populations of Variable Size. *Genetics*, 186(2):601–611, Oct. 2010. ISSN 0016-6731. doi: 10.1534/genetics.110.117135. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2954466/.

[14] F. D. Farrell, M. Gralka, O. Hallatschek, and B. Waclaw. Mechanical interactions in bacterial colonies and the surfing probability of beneficial mutations. *Journal of The Royal Society Interface*, 14(131):20170073, June 2017. doi: 10.1098/rsif.2017.0073. URL https://royalsocietypublishing.org/doi/full/10.1098/rsif.2017.0073.

[15] G. Francius, P. Polyakov, J. Merlin, Y. Abe, J.-M. Ghigo, C. Merlin, C. Beloin, and J. F. L. Duval. Bacterial Surface Appendages Strongly Impact Nanomechanical and Electrokinetic Properties of Escherichia coli Cells Subjected to Osmotic Stress. *PLoS ONE*, 6(5):e20066, May 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0020066. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0020066.

[16] P. D. Freese, K. S. Korolev, J. I. Jiménez, and I. A. Chen. Genetic drift suppresses bacterial conjugation in spatially structured populations. *Biophysical Journal*, 106(4): 944–954, 2014.

[17] S. French, J.-P. Côté, J. M. Stokes, R. Truant, and E. D. Brown. Bacteria Getting into Shape: Genetic Determinants of E. coli Morphology. *mBio*, 8(2), May 2017. ISSN 2150-7511. doi: 10.1128/mBio.01977-16. URL https://mbio.asm.org/content/8/2/e01977-16.

[18] S. D. W. Frost, M.-J. Dumaurier, S. Wain-Hobson, and A. J. L. Brown. Genetic drift and within-host metapopulation dynamics of HIV-1 infection. *Proceedings of the National Academy of Sciences*, 98(12):6975–6980, June 2001. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.131056998. URL https://www.pnas.org/content/98/12/6975.

[19] D. Fusco, M. Gralka, J. Kayser, A. Anderson, and O. Hallatschek. Excess of mutational jackpot events in expanding populations revealed by spatial Luria–Delbrück experiments. *Nature Communications*, 7(1):1–9, Oct. 2016. ISSN 2041-1723. doi: 10.1038/ncomms12760. URL https://www.nature.com/articles/ncomms12760.

[20] D. R. Gifford, J. A. G. M. de Visser, and L. M. Wahl. Model and test in a fungus of the probability that beneficial mutations survive drift. *Biology Letters*, 9(1):20120310, Feb. 2013. doi: 10.1098/rsbl.2012.0310. URL https://royalsocietypublishing.org/doi/full/10.1098/rsbl.2012.0310. Publisher: Royal Society.

[21] J. H. Gillespie. *Population Genetics: A Concise Guide*. JHU Press, Aug. 2004. ISBN 978-1-4214-0170-6. Google-Books-ID: KAcAfiyHpcoC.

[22] D. S. Glass and I. H. Riedel-Kruse. A synthetic bacterial cell-cell adhesion toolbox for programming multicellular morphologies and patterns. *Cell*, 174(3):649–658, 2018.

[23] F. Goldschmidt, R. R. Regoes, and D. R. Johnson. Metabolite toxicity slows local diversity loss during expansion of a microbial cross-feeding community. *The ISME journal*, 12(1):136–144, 2018.

[24] B. H. Good and M. M. Desai. Evolution of Mutation Rates in Rapidly Adapting Asexual Populations. *Genetics*, 204(3):1249–1266, Nov. 2016. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.116.193565. URL https://www.genetics.org/content/204/3/1249. Publisher: Genetics Section: Investigations.

[25] M. Gralka and O. Hallatschek. Environmental heterogeneity can tip the population genetics of range expansions. *eLife*, 8:e44359, Apr. 2019. ISSN 2050-084X. doi: 10.7554/eLife.44359. URL https://doi.org/10.7554/eLife.44359. Publisher: eLife Sciences Publications, Ltd.

[26] M. Gralka, F. Stiewe, F. Farrell, W. Möbius, B. Waclaw, and O. Hallatschek. Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters*, 19(8):889–898, 2016. ISSN 1461-0248. doi: 10.1111/ele.12625. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.12625.

[27] O. Hallatschek and D. R. Nelson. Life at the Front of an Expanding Population. *Evolution*, 64(1):193–206, 2010. ISSN 1558-5646. doi: 10.1111/j.1558-5646.2009.00809.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.2009.00809.x.

[28] O. Hallatschek, P. Hersen, S. Ramanathan, and D. R. Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930, Dec. 2007. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0710150104. URL https://www.pnas.org/content/104/50/19926.

[29] A. Harpak and G. Sella. Neutral null models for diversity in serial transfer evolution experiments. *Evolution*, 68(9):2727–2736, 2014.

[30] D. L. Hartl and A. G. Clark. *Principles of Population Genetics*. Sinauer Associates, 1989. ISBN 978-0-87893-302-0.

[31] M. v. d. Heijden, D. M. Miedema, B. Waclaw, V. L. Veenstra, M. C. Lecca, L. E. Nijman, E. v. Dijk, S. M. v. Neerven, S. C. Lodestijn, K. J. Lenos, N. E. d. Groot, P. R. Prasetyanti, A. A. Varea, D. J. Winton, J. P. Medema, E. Morrissey, B. Ylstra, M. A. Nowak, M. F. Bijlsma, and L. Vermeulen. Spatiotemporal regulation of clonogenicity

in colorectal cancer xenografts. *Proceedings of the National Academy of Sciences*, 116 (13):6140–6145, Mar. 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1813417116. URL https://www.pnas.org/content/116/13/6140.

[32] A. Kan, I. Del Valle, T. Rudge, F. Federici, and J. Haseloff. Intercellular adhesion promotes clonal mixing in growing bacterial populations. *Journal of The Royal Society Interface*, 15(146):20180406, Sept. 2018. doi: 10.1098/rsif.2018.0406. URL https://royalsocietypublishing.org/doi/10.1098/rsif.2018.0406.

[33] J. Kayser, C. F. Schreck, Q. Yu, M. Gralka, and O. Hallatschek. Emergence of evolutionary driving forces in pattern-forming microbial populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1747):20170106, 2018.

[34] M. Kimura. Evolutionary Rate at the Molecular Level. *Nature*, 217(5129):624–626, Feb. 1968. ISSN 1476-4687. doi: 10.1038/217624a0. URL https://www.nature.com/articles/217624a0.

[35] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1983. ISBN 978-0-521-31793-1.

[36] J. L. King and T. H. Jukes. Non-Darwinian Evolution. *Science*, 164(3881):788–798, May 1969. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.164.3881.788. URL https://science.sciencemag.org/content/164/3881/788.

[37] K. S. Korolev, M. J. I. Müller, N. Karahan, A. W. Murray, O. Hallatschek, and D. R. Nelson. Selective sweeps in growing microbial colonies. *Physical Biology*, 9(2):026008, Apr. 2012. ISSN 1478-3975. doi: 10.1088/1478-3975/9/2/026008. URL https://doi.org/10.1088%2F1478-3975%2F9%2F2%2F026008.

[38] T. Kruse, J. Møller-Jensen, A. Løbner-Olesen, and K. Gerdes. Dysfunctional MreB inhibits chromosome segregation in Escherichia coli. *The EMBO journal*, 22(19):5283–5292, 2003.

[39] M. O. Lavrentovich, K. S. Korolev, and D. R. Nelson. Radial Domany-Kinzel models with mutation and selection. *Physical Review E*, 87(1):012103, 2013.

[40] S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, and G. Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519 (7542):181–186, Mar. 2015. ISSN 1476-4687. doi: 10.1038/nature14279. URL https://www.nature.com/articles/nature14279.

[41] J. Lin and A. Amir. From single-cell variability to population growth. *Physical Review E*, 101(1):012401, Jan. 2020. doi: 10.1103/PhysRevE.101.012401. URL https://link.aps.org/doi/10.1103/PhysRevE.101.012401.

[42] C. Lopez-Garcia, A. M. Klein, B. D. Simons, and D. J. Winton. Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift. *Science*, 330(6005):822–825, Nov. 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1196236. URL https://science.sciencemag.org/content/330/6005/822.

[43] M. Lynch. Evolution of the mutation rate. *Trends in Genetics*, 26(8):345–352, Aug. 2010. ISSN 0168-9525. doi: 10.1016/j.tig.2010.05.003. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910838/.

[44] M. Lynch, B. Koskella, and S. Schaack. Mutation Pressure and the Evolution of Organelle Genomic Architecture. *Science*, 311(5768):1727–1730, Mar. 2006. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1118884. URL https://science.sciencemag.org/content/311/5768/1727.

[45] J. T. McCrone, R. J. Woods, E. T. Martin, R. E. Malosh, A. S. Monto, and A. S. Lauring. Stochastic processes constrain the within and between host evolution of influenza virus. *eLife*, 7:e35962, Apr. 2018. ISSN 2050-084X. doi: 10.7554/eLife.35962. URL https://doi.org/10.7554/eLife.35962.

[46] S. Mitri, E. Clarke, and K. R. Foster. Resource limitation drives spatial organization in microbial groups. *The ISME Journal*, 10(6):1471–1482, June 2016. ISSN 1751-7370. doi: 10.1038/ismej.2015.208. URL https://www.nature.com/articles/ismej2015208.

[47] L. D. Mueller, L. G. Barr, and F. J. Ayala. Natural selection vs. random drift: evidence from temporal variation in allele frequencies in nature. *Genetics*, 111(3):517–554, 1985.

[48] M. J. Müller, B. I. Neugeboren, D. R. Nelson, and A. W. Murray. Genetic drift opposes mutualism during spatial population expansion. *Proceedings of the National Academy of Sciences*, 111(3):1037–1042, 2014.

[49] C. D. Nadell, K. R. Foster, and J. B. Xavier. Emergence of Spatial Structure in Cell Groups and the Evolution of Cooperation. *PLoS Computational Biology*, 6(3):e1000716, Mar. 2010. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000716. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000716.

[50] E. T. E. Niba, Y. Naka, M. Nagase, H. Mori, and M. Kitakawa. A Genome-wide Approach to Identify the Genes Involved in Biofilm Formation in E. coli. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 14(6):237–246, 2007. ISSN 1340-2838. doi: 10.1093/dnares/dsm024. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2779908/.

[51] T. Ohta. Slightly Deleterious Mutant Substitutions in Evolution. *Nature*, 246(5428):96–98, Nov. 1973. ISSN 1476-4687. doi: 10.1038/246096a0. URL https://www.nature.com/articles/246096a0.

[52] E. R. Oldewurtel, N. Kouzel, L. Dewenter, K. Henseler, and B. Maier. Differential interaction forces govern bacterial sorting in early biofilms. *eLife*, 4:e10811, Sept. 2015. ISSN 2050-084X. doi: 10.7554/eLife.10811. URL https://doi.org/10.7554/eLife.10811. Publisher: eLife Sciences Publications, Ltd.

[53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. ISSN 1533-7928. URL http://jmlr.org/papers/v12/pedregosa11a.html.

[54] C. R. Reisch and K. L. J. Prather. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. *Scientific Reports*, 5(1):15096, Oct. 2015. ISSN 2045-2322. doi: 10.1038/srep15096. URL https://www.nature.com/articles/srep15096. Number: 1 Publisher: Nature Publishing Group.

[55] S. Sacristán, J. M. Malpica, A. Fraile, and F. García-Arenal. Estimation of Population Bottlenecks during Systemic Movement of Tobacco Mosaic Virus in Tobacco Plants. *Journal of Virology*, 77(18):9906–9911, Sept. 2003. ISSN 0022-538X, 1098-5514. doi: 10.1128/JVI.77.18.9906-9911.2003. URL https://jvi.asm.org/content/77/18/9906.

[56] M. Schmutzer and A. Wagner. Gene expression noise can promote the fixation of beneficial mutations in fluctuating environments. *bioRxiv*, page 2020.02.13.947275, Feb. 2020. doi: 10.1101/2020.02.13.947275. URL https://www.biorxiv.org/content/10.1101/2020.02.13.947275v1.

[57] H. Shi, A. Colavin, M. Bigos, C. Tropini, R. D. Monds, and K. C. Huang. Deep Phenotypic Mapping of Bacterial Cytoskeletal Mutants Reveals Physiological Robustness to Cell Size. *Current Biology*, 27(22):3419–3429.e4, Nov. 2017. ISSN 0960-9822. doi: 10.1016/j.cub.2017.09.065. URL http://www.sciencedirect.com/science/article/pii/S0960982217312654.

[58] W. P. J. Smith, Y. Davit, J. M. Osborne, W. Kim, K. R. Foster, and J. M. Pitt-Francis. Cell morphology drives spatial patterning in microbial communities. *Proceedings of the National Academy of Sciences*, 114(3):E280–E286, Jan. 2017. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1613007114. URL https://www.pnas.org/content/114/3/E280.

[59] H. P. Steenackers, I. Parijs, K. R. Foster, and J. Vanderleyden. Experimental evolution in biofilm populations. *FEMS Microbiology Reviews*, 40(3):373–397, May 2016. ISSN 0168-6445. doi: 10.1093/femsre/fuw002. URL https://academic.oup.com/femsre/article/40/3/373/2570137.

[60] R. Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 1996. ISSN 2517-6161. doi:

10.1111/j.2517-6161.1996.tb02080.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x.

[61] T. Ursell, T. K. Lee, D. Shiomi, H. Shi, C. Tropini, R. D. Monds, A. Colavin, G. Billings, I. Bhaya-Grossman, M. Broxton, B. E. Huang, H. Niki, and K. C. Huang. Rapid, precise quantification of bacterial cellular dimensions across a genomic-scale knockout library. *BMC Biology*, 15(1):17, Feb. 2017. ISSN 1741-7007. doi: 10.1186/s12915-017-0348-8. URL https://doi.org/10.1186/s12915-017-0348-8.

[62] J. van Gestel, M. Ackermann, and A. Wagner. Microbial life cycles link global modularity in regulation to mosaic evolution. *Nature Ecology & Evolution*, 3(8): 1184–1196, Aug. 2019. ISSN 2397-334X. doi: 10.1038/s41559-019-0939-6. URL https://www.nature.com/articles/s41559-019-0939-6.

[63] M. J. Williams, B. Werner, C. P. Barnes, T. A. Graham, and A. Sottoriva. Identification of neutral tumor evolution across cancer types. *Nature Genetics*, 48(3):238–244, Mar. 2016. ISSN 1546-1718. doi: 10.1038/ng.3489. URL https://www.nature.com/articles/ng.3489.

[64] E. S. Wright and K. H. Vetsigian. Stochastic exits from dormancy give rise to heavy-tailed distributions of descendants in bacterial populations. *Molecular Ecology*, 28(17):3915–3928, 2019. ISSN 1365-294X. doi: 10.1111/mec.15200. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15200.

[65] A. A. Zaidi, P. R. Wilton, M. S.-W. Su, I. M. Paul, B. Arbeithuber, K. Anthony, A. Nekrutenko, R. Nielsen, and K. D. Makova. Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees. *Proceedings of the National Academy of Sciences*, 116(50):25172–25178, Dec. 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1906331116. URL https://www.pnas.org/content/116/50/25172.

Chapter 3

Lineage frequency time series reveal elevated levels of genetic drift in SARS-CoV-2 transmission in England

Genetic drift is also an important aspect of pathogen transmission, where a "birth" occurs when the pathogen is transmitted between two hosts, and a "death" occurs when an infected host recovers or dies. Here, we address a major challenge in inferring the strength of genetic drift in pathogen transmission in the natural world, which is how to account for measurement or observation noise [23]. We develop a new method to infer genetic drift simultaneously with measurement noise using lineage frequency time series data. We apply the method to study the strength of genetic drift in SARS-CoV-2 transmission in England.

I am grateful to Joao Ascensao for developing the statistical inference method, Takashi Okada for creating lineages from phylogenetic trees and drafting the respective method section, Olivia Boyd for sourcing the trees and metadata from COG-UK, Oskar Hallatschek for advising, and all coauthors for giving helpful feedback on the work and writing presented in this chapter. The remainder of this chapter will be submitted as QinQin Yu, Joao Ascensao, Takashi Okada, Olivia Boyd, Erik Volz, Oskar Hallatschek, Lineage frequency time series reveal elevated levels of genetic drift in SARS-CoV-2 transmission in England, 2022.

3.1 Abstract

Random genetic drift in the population-level dynamics of infectious disease results from the randomness of inter-host transmission and the randomness of host recovery. The strength of genetic drift has been found to be high for SARS-CoV-2 due to superspreading, and this is expected to substantially impact the disease epidemiology and evolution. Noise that results from the measurement process, such as biases in data collection across time, geographical areas, etc., can potentially confound estimates of genetic drift as both processes contribute "noise" to the data. To address this challenge, we develop a method to jointly infer genetic

drift and measurement noise from time-series lineage frequency data. We apply this method to SARS-CoV-2 genomic data from England from March 2020 until December 2021. We find that even after correcting for measurement noise, the strength of genetic drift is consistently, throughout time, higher than that expected from the observed number of COVID-19 positive individuals in England by 1 to 3 orders of magnitude. Corrections taking into account susceptible-exposed-infected-recovered epidemiological dynamics do not explain the discrepancy. The levels of genetic drift that we observe are higher than the estimated levels of superspreading found by modeling studies that incorporate data on actual contact statistics in England. We discuss how even in the absence of superspreading, high levels of genetic drift can be generated via jackpot events in a deme model. Our results suggest that understanding heterogeneous host contact structure may be important for understanding the high levels of genetic drift observed for SARS-CoV-2 in England.

3.2 Introduction

Random genetic drift is the change in the composition of a population over time due to the randomness of birth and death processes. In pathogen transmission, births occur as a result of transmission of the pathogen between hosts and deaths occur as a result of infected host recovery or death. The strength of genetic drift in pathogen transmission is set by the disease prevalence, the disease epidemiology parameters [64], the variance in offspring number (the number of secondary infections that result from an infected individual) [35], as well as host contact patterns [3]. Many diseases have been found to exhibit high levels of genetic drift, such as SARS, MERS, tuberculosis, and measles [35, 31, 40]. The strength of genetic drift affects how the disease spreads through the population [35, 3, 46] how new variants emerge [44, 25, 14, 59], and the effectiveness of interventions [54], making it an important quantity to accurately estimate for understanding disease epidemiology, evolution, and control. The effective population size is often used to quantify the strength of genetic drift; it is the population size in an idealized Wright-Fisher model (with discrete non-overlapping generations, a constant population size, and offspring determined by sampling with replacement from the previous generation) that would reproduce the observed dynamics [11]. If the effective population size is lower than the true population size, it is an indication that there are additional sources of stochasticity beyond the Wright-Fisher model; thus, a lower effective population size indicates a higher level of genetic drift.

Data from contact tracing and phylogenetics have revealed high levels of superspreading and genetic drift for SARS-CoV-2 [32, 4, 24]. However, it is extremeley tedious to estimate the strength of genetic drift as it requires an unbiased dataset, and most studies have focused on a particular time period or geographical location (see phylogenetic studies [63, 49, 17], and see Supplementary table 3.S1 for non-phylogenetic studies); we still lack a systematic study of genetic drift over time and across different geographic subdivisions. It is also challenging to accurately estimate the strength of genetic drift because non-biological factors, such as the data collection process, also contribute noise to the dataset [23]. Observation, or

measurement, noise can arise from a variety of factors, including variability in the testing rate across time, geographic locations, demographic groups, and symptom status, and biases in contact tracing. Some phylogenetic and time-series frequency methods that infer the strength of genetic drift account for uniform sampling of sequences from the population [56, 7, 8], but in reality, the observation process may contribute more variance than purely from uniform sampling due to the complex decisions that lead to which infected individuals end up being sampled. A recent study has accounted for overdispersed sampling of sequences in the inference of fitness coefficients of SARS-CoV-2 variants, but assumes constant overdispersion over time [16]; in reality, the observation process may change over time. One approach is to try to collect as unbiased of a dataset as possible, but this may not always be possible, for instance when resources are scarce. In this study we consider the question of what signatures measurement noise might leave, particularly in time-series frequency data, and if and how we might be able to infer its strength to decouple it from genetic drift.

Inference of measurement noise can be incorporated straightforwardly into time-series lineage or allele frequency methods [7, 8] (see the Supplementary information for a summary of other methods used for inferring genetic drift, and additional references). In these methods, time series data of the abundance of different lineages or alleles (groups of samples that are genetically or phenotypically similar) are fit to models that incorporate genetic drift and measurement noise. Intuitively, in time-series frequency data, genetic drift leads to frequency fluctuations whose magnitudes scale with time, whereas measurement noise leads to frequency fluctuations whose magnitudes do not scale with time (Figure 3.1a). Mathematically, this problem maps well to a Hidden Markov Model with continuous hidden and observed states (similar to a Kalman filter), where the hidden states are the true frequencies and the observed states are the observed frequencies (Figure 3.1b), and the processes of genetic drift and measurement noise determine the transition and emission probabilities, respectively [66, 57]. Here we develop a method to jointly infer genetic drift and measurement noise that allows measurement noise to be overdispersed compared to uniform sampling and for the strength of overdispersion to vary over time, which is an improvement from previous time-series methods. By fitting the model to the observed lineage frequency trajectories, we show that the effective population size and the strength of measurement noise can be accurately determined in most situations, even when both quantities are varying over time.

We apply this method to study the strengths of genetic drift and measurement noise for SARS-CoV-2 in England from March 2020 until December 2021. We focus on England due to its consistently large number of sequenced SARS-CoV-2 cases since early in the pandemic, which allows us to most accurately estimate lineage frequency time series. We find that even after correcting for measurement noise, the strength of genetic drift is consistently, throughout time, higher than expected from the observed number of infected individuals in England by 1 to 3 orders of magnitude. Corrections taking into account epidemiological dynamics, such as a susceptible-exposed-infected-recovered model, does not explain the discrepancy. The levels of genetic drift that we observe are higher than the estimated levels of superspreading found by modeling studies that incorporate data on actual contact statistics in England. We discuss how even in the absence of superspreading, these high levels of

genetic drift can be generated via jackpot events in a deme model. Below, we first present the method and validation, and then the application to SARS-CoV-2 in England.

3.3 Results

Method for jointly inferring genetic drift and measurement noise from time-series lineage frequency data

We first summarize the statistical inference method that we developed to infer time-varying effective population sizes from neutral lineage frequency time series that are affected by overdispersed measurement noise (more variable than uniform sampling). We explain the method more extensively in the Methods. Briefly, we use a Hidden Markov Model (HMM) with continuous hidden and observed states (similar to a Kalman filter), where the hidden states are the true frequencies (f_t, where t is time), and the observed states are the observed frequencies (f_t^{obs}) (Figure 3.1b) (see Methods). The transition probability between hidden states is set by genetic drift, where the mean true frequency is the true frequency at the previous time $\mathrm{E}(f_{t+1}|f_t) = f_t$, and when the frequencies are rare the variance in frequency is proportional to the mean, $\mathrm{Var}(f_{t+1}|f_t) = \frac{f_t}{\tilde{N}_e(t)}$. $\tilde{N}_e(t) = N_e(t)\tau(t)$ where $N_e(t)$ is the effective population size and $\tau(t)$ is the generation time, and both quantities can vary over time; however, we are only able to infer the compound parameter $N_e(t)\tau(t)$. The emission probability between hidden and observed states is set by measurement noise, where the mean observed frequency is the true frequency $\mathrm{E}(f_t^{obs}|f_t) = f_t$ and when the frequencies are rare the variance in the observed frequency is proportional to the mean, $\mathrm{Var}(f_t^{obs}|f_t) = c_t\frac{f_t}{M_t}$. $c_t \geq 1$ describes the deviation from uniform sampling ($c_t = 1$), and M_t is the number of sequences at time t. Our model assumes that the number of individuals and frequency of a lineage is high enough such that the central limit theorem applies; to meet this condition, we created "superlineages" where we randomly and exclusively grouped lineages together such that the sum of their abundances and frequencies was above a threshold (see Methods).

Using the transition and emission probability distributions (see Methods) and the HMM structure, we determine the likelihood function (Equation 3.13 in Methods) describing the probability of observing a particular set of lineage frequency time-series data given the unknown parameters, namely the scaled effective population size across time $\tilde{N}_e(t)$ and the strength of measurement noise across time c_t. We then maximize the likelihood over the parameters to determine the most likely parameters that describe the data. Because we are relying on a time-series signature in the data for the inference, we need to use a sufficiently large number of timesteps of data, but on the other hand, the longer the time series, the more parameters would need to be inferred (since both $\tilde{N}_e(t)$ and c_t are allowed to change over time). To balance these two factors, we assumed that the effective population size stays constant over a time period of 9 weeks (a form of "regularization"). We then shift this window of 9 weeks across time to determine how $\tilde{N}_e(t)$ changes over time (see Methods), but this effectively averages the inferred $\tilde{N}_e(t)$ over time. c_t is still allowed to vary weekly.

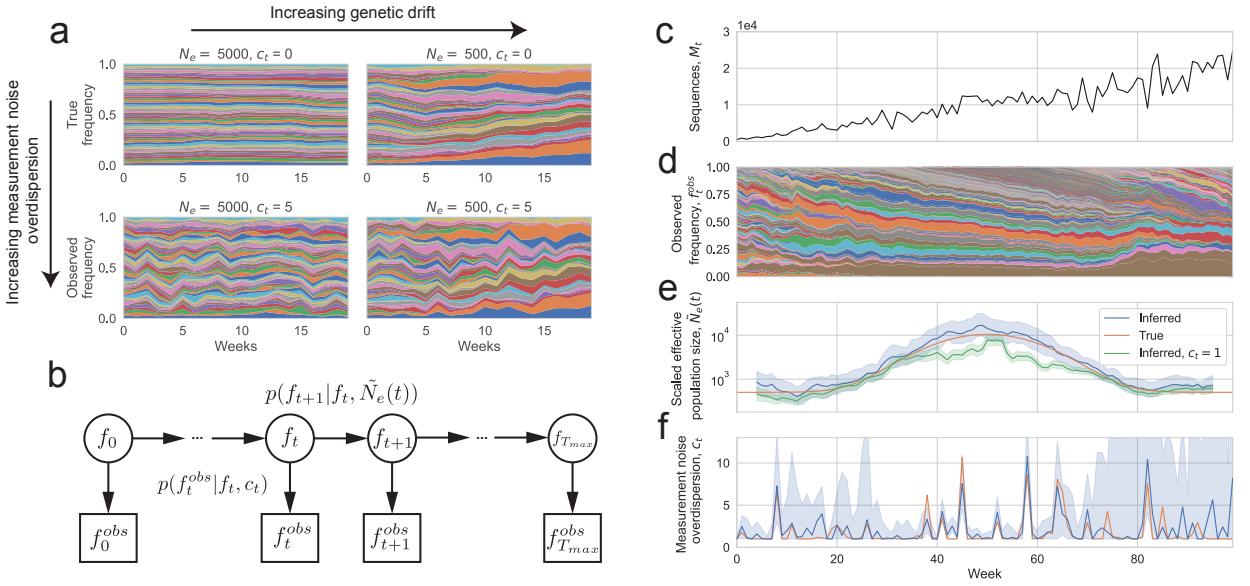

Figure 3.1: A Hidden Markov Model with continuous hidden and observed states (similar to a Kalman filter) for inferring genetic drift and measurement noise from lineage frequency time series. (a) Illustration of how genetic drift and measurement noise affect the observed frequency time series. Muller plot of lineage frequencies from Wright-Fisher simulations with effective population size 500 and 5000, with and without measurement noise. In simulations with measurement noise, 100 sequences were sampled per week with the measurement noise overdispersion parameter $c_t = 5$, parameter defined in text). All simulations were initialized with 50 lineages at equal frequency. A lower effective population size leads to larger frequency fluctuations whose variances add over time, whereas measurement noise leads to increased frequency fluctuations whose variances do not add over time. (b) Schematic of Hidden Markov Model describing frequency trajectories. f_t is the true frequency at time t (hidden states) and f_t^{obs} is the observed frequency at time t (observed states). The inferred parameters are $\tilde{N}_e(t) \equiv N_e(t)\tau(t)$, the effective population size scaled by the generation time, and c_t, the overdispersion in measurement noise ($c_t = 1$ corresponds to uniform sampling of sequences from the population). (c-f) Validation of method using Wright-Fisher simulations of frequency trajectories with time-varying effective population size and measurement noise. (c) Simulated number of sequences. (d) Simulated lineage frequency trajectories. (e) Inferred scaled effective population size ($\tilde{N}_e(t)$) on simulated data compared to true values. (f) Inferred measurement noise (c_t) on simulated data compared to true values. In (e) the shaded region shows the 95% confidence interval calculated using the posterior, and in (f) the shaded region shows the 95% confidence interval calculated using bootstrapping (see Methods).

To validate our model, we ran Wright-Fisher simulations with time-varying effective population size and time-varying measurement noise (Figure 3.1c-f). Because a substantial number of lineages would go extinct over the simulation timescale of 100 weeks, we introduced mutations with a small rate ($\mu = 0.01$ per week per individual) to prevent the number of lineages from becoming too low. We then did inference on the simulated time-series frequency trajectories (Figure 3.1d). The inferred $\tilde{N}_e(t)$ and c_t closely follow the true values (Figure 3.1e-f), and the 95% confidence intervals (see Methods for how they are calculated) include the true value in a median (across timepoints) of 95% of simulation realizations (Figure 3.S2). The error in c_t is higher when the variance contributed to the frequency trajectories by measurement noise is lower than that of genetic drift, which occurs when the effective population size is low or number of sequences is high (more clearly seen in Figure 3.S4, where the effective population size is held constant). However, the error on $\tilde{N}_e(t)$ seems to be unchanged or even slightly decrease when the error on c_t is increased because the contribution to the variance due to genetic drift is higher. We also observe that the inferred $\tilde{N}_e(t)$ is smoothed over time due to the assumption of constant $\tilde{N}_e(t)$ over 9 weeks (Figure 3.S3); this is a potential drawback when there are sharp changes in the effective population size over time. Importantly, we observed that the inferred $\tilde{N}_e(t)$ will be underestimated if sampling is assumed to be uniform when it is actually overdispersed (Figure 3.1e). This is because variance in the frequency trajectories due to measurement noise is incorrectly being attributed to genetic drift. The underestimation is strongest when the variance contributed due to measurement noise is high, either due to high measurement noise overdispersion, a low number of sampled sequences, or a high effective population size. In this situation, joint inference of measurement noise and $\tilde{N}_e(t)$ from the data is necessary for accurate inference of $\tilde{N}_e(t)$.

Application to COG-UK data in England

We next applied this method to study the effective population size and strength of measurement noise for SARS-CoV-2 in England. Because our method assumes that lineages are neutral with respect to one another (no selection), we performed separate analyses on groups of lineages that have been shown to exhibit fitness differences or deterministic changes in frequency: lineages pre-B.1.177, B.1.177, Alpha, and Delta [16, 63, 10, 13]. We did not find any studies in the literature claiming detectable fitness differences between lineages within each of these groups; thus, we assumed that our neutral model should be valid when analyzing lineages only within a single group.

To obtain lineage frequency time series data for SARS-CoV-2 in England, we downloaded genomic metadata from the COVID-19 Genomics UK Consortium (COG-UK) [60] (Figure 3.2b) and the associated phylogenetic trees that were created at different points in time. For sequences pre-B.1.177, we used the pangolin lineages assignments from COG-UK [45, 50]. However, B.1.177, Alpha, and Delta were subdivided into few or only one pangolin lineage, since a new lineage is defined by sufficiently many mutations and evidence of geographic importation. However, these requirements are not important for our purposes

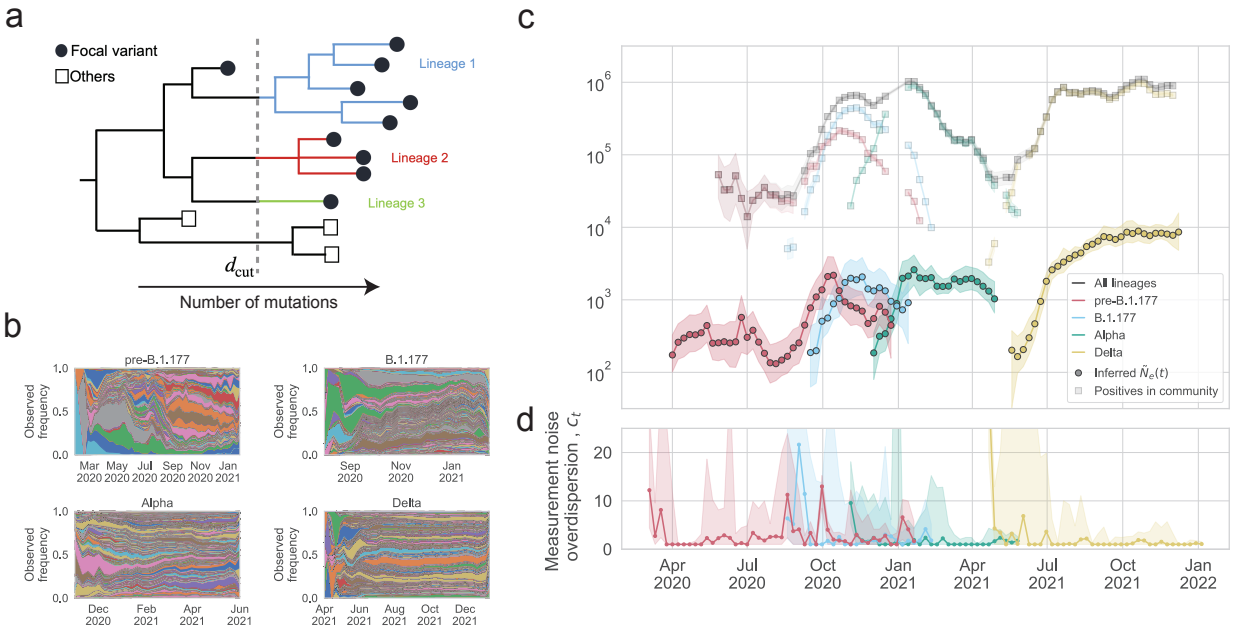

Figure 3.2: The inferred effective population size and overdispersion of measurement noise in England compared with the number of positive individuals. (a) Schematic of lineage construction for B.1.177, Alpha, and Delta from the COG-UK phylogenetic tree. The filled circles represent the sequences of a focal variant sampled in England, while the unfilled squares represent other sequences, which are of other variants or sampled in other countries. The phylogenetic tree is cut at a certain depth $d = d_{\text{cut}}$, and each branch cut by the line $d = d_{\text{cut}}$ defines a lineage. Lineages pre-B.1.1.7 are defined using the pango nomenclature [45, 50]. (b) Muller plot of lineage frequency time series for lineages pre-B.1.177, of B.1.177, of Alpha, of Delta. (c) Inferred scaled effective population size ($\tilde{N}_e(t) \equiv N_e(t)\tau(t)$) for pre-B.1.177 sequences, B.1.177, Alpha, and Delta, compared to the estimated number of people testing positive for SARS-CoV-2 in England at the community level, as measured by the COVID-19 Infection Survey [62], for all lineages and by variant or group of lineages. To simplify the plot, only data where the number of positive individuals for a given variant or group of lineages was higher than 10^3 in a week are shown. The inferred $\tilde{N}_e(t)$ is considerably lower than the number of positive individuals for all times and for all variants or group of lineages. (d) Inferred measurement noise overdispersion (c_t) for pre-B.1.177 sequences, B.1.177, Alpha, and Delta.

and instead we only need resolution of neutral lineages within a variant. Thus, we created our own neutral lineages by grouping sequences together based on phylogenetic distance in the tree (see Figure 3.2a and Methods), and cutting the tree at a particular point. Most sequenced samples were included in the trees (Figure 3.S6), and any downsampling was done

by preserving genetic diversity. Most sequences in the tree were assigned to lineages (see Methods), and we corrected for the fraction of sequences that were not assigned to lineages in our inference of $\tilde{N}_e(t)$ (see Methods). This yielded 486 lineages for pre-B.1.177, 4083 lineages for B.1.177, 6225 lineages for Alpha, 24867 lineages for Delta.

The inferred effective population size is shown in Figure 3.2c. The inferred effective population size was lower than the number of positive individuals in the community by a factor of 30 to 1070 at different points in time, but closely followed the shape of the number of positive individuals with a few exceptions: the inferred effective population size of lineages pre-B.1.177 peaked slightly before the number of pre-B.1.177 positives peaked, the inferred effective population size of Alpha stayed relatively constant after January 2021 while the number of positives decreased, and the shoulder for the inferred effective population size of Delta occurred earlier than in the number of positives.

The inferred measurement noise for each group of lineages is shown in Figure 3.2d. In summary, the inferred measurement noise overdispersion was mostly indistinguishable from 1 (uniform sampling), but at times was above 1 (sampling that is more variable than uniform sampling), particularly when a variant was first emerging. There were also at times differences in the strength of measurement noise between variants when they overlapped in time. Measurement noise for lineages pre-B.1.177 was high at the beginning of the pandemic in March 2020, then became indistinguishable from uniform sampling from April 2020 to January 2021, except for a sharp peak in October 2020. Measurement noise for B.1.177 was generally indistinguishable from uniform sampling. The measurement noise of Alpha was high when it first emerged (November 2020) despite sampling noise being low for other lineages that were circulating at that time. The measurement noise of Delta was high when it first emerged (April to May 2021), but dropped to around 1 after July 2021 and stayed low.

To better understand the observed levels of genetic drift, we compared the inferred $\tilde{N}_e(t)$ to that of a null model. We chose an SEIR model, which includes a susceptible, exposed, infectious, and recovered class. The SEIR model is a good representation of the epidemiology of SARS-CoV-2 when the infectious class includes both asymptomatic and symptomatic individuals, i.e. the exposed class describes pre-symptomatic individuals who ultimately become positive no matter whether they become symptomatic or not. In our case, we have data on the number of positives that includes both symptomatic and asymptomatic individuals, so we have the correct input data for this model. We derived the $\tilde{N}_e(t)$ for an SEIR model (see Methods):

$$\tilde{N_e}^{\text{SEIR}}(t) \equiv \{N_e(t)\tau(t)\}^{\text{SEIR}} = \frac{(E(t)+I(t))^2}{2R_t\gamma_I I(t)}. \tag{3.1}$$

where $E(t)$ is the number of exposed individuals, $I(t)$ is the number of infectious individuals, R_t is the effective reproduction number, and γ_I is the rate at which infectious individuals stop being infectious. For the number of infectious individuals, we used the number of positive individuals estimated from the UK Office for National Statistics' COVID-19 Infection

Survey [62], which is a household surveillance study that reports positive PCR tests, regardless of symptom status. We used the measured effective reproduction number in England reported by the UK Health Security Agency [61]. We found that $\tilde{N_e}^{\text{SEIR}}(t)$ is very similar to the number of positives because the effective reproduction number in England was very close to 1 across time. To calculate $\tilde{N_e}^{\text{SEIR}}(t)$ for each variant or group of lineages, we rescaled the population-level $I(t)$ and R_t based on the fraction of each variant in the population and the relative differences in reproduction numbers between variants (see Methods). We then calculated the scaled true population size, $\tilde{N}(t) \equiv N(t)\tau(t)$, for the SEIR model by multiplying by the variance in offspring number, σ^2, for the SEIR model [30]

$$\tilde{N}^{\text{SEIR}}(t) = \tilde{N_e}^{\text{SEIR}}(t)\{\sigma^2\}^{\text{SEIR}} \tag{3.2}$$

$$\{\sigma^2\}^{\text{SEIR}} = 2. \tag{3.3}$$

Overall, the inferred $\tilde{N_e}(t)$ is lower than $\tilde{N}^{\text{SEIR}}(t)$ by a time-dependent factor that varies between 100 and 2350 (Figures 3.3c and 3.S9), suggesting high levels of genetic drift in England across time. The ratio of $\tilde{N}^{\text{SEIR}}(t)$ to the inferred $\tilde{N_e}(t)$ was higher for Alpha than for Delta, suggesting that the level of genetic drift for Alpha was higher than expected, when compared to that of Delta.

We also probed the spatial structure of transmission by inferring the scaled effective population size separately for each region within England. We find that the scaled effective population size in the regions of England is substantially smaller than that in England as a whole for Alpha and Delta (Figure 3.S1), suggesting that the transmission was not well-mixed at that time. Additionally, the discrepancy between the inferred regional scaled effective population size and the observed number of positive individuals in a region was comparable to that seen in England as a whole (Figure 3.S10), which is consistent with spatially segregated dynamics with similar levels of genetic drift in each region. We further describe these results in the Supplementary Information.

Potential mechanisms that can contribute to the high levels of genetic drift

Two potential mechanisms that can contribute to the observed high levels of genetic drift are: (1) variability at the individual level through superspreading (Figure 3.3a), and (2) host population structure (Figure 3.3b). We investigate each of these mechanisms in turn and compare it to our results. While in reality, both mechanisms (and others not explored here) are likely at play, it is challenging to tease them apart given our limited data. Instead, we consider the extreme situations where one or the other mechanism is driving the dynamics to gain intuition.

Superspreading occurs due to overdispersion in the number of secondary cases, which decreases the effective population size. If superspreading were the only mechanism at play, then the variance in offspring number that would explain our results would be the same as the

ratio between the SEIR null model $\tilde{N}^{\text{SEIR}}(t)$ and the inferred $\tilde{N}_e(t)$ (100-2350) (Figure 3.3c). Current estimates of the variance in offspring number measured by contact tracing and modeling across a wide range of times and locations are from around 0.7 in one study to 65 in another (Table 3.S1). We found two studies that apply to the UK, one which used a model that incorporated the empirical viral load trajectories and contact numbers to estimate superspreading [48] and another which used a branching process model of the number of imported and local cases [19]. The only of those estimates whose time window overlapped with our time windows found substantially lower levels of superspreading than what we observe (Figure 3.3c). It is possible that contact tracing and modeling over- or under-estimates overdispersion due to missed contacts. However, on the other hand, it may be the case that superspreading is not the only mechanism at play.

We propose another mechanism that can lead to a decreased effective population size: host deme structure. In such a model, individuals within a deme are very well-connected to one another (i.e. households or friend groups, also known as "communities" in network science [42]), but there are few connections between demes (Figure 3.3c). It is possible for deme structure to occur without superspreading. For instance, in the schematic in Figure 3.3c, the number of contacts is either 4 or 5; if every contact led to a transmission, this would be an extremely narrow offspring number distribution (i.e. no superspreading). Because individuals are very well-connected within a deme, once the pathogen spreads to a susceptible deme, it will spread rapidly in a deme until all individuals are infected (a jackpot event). In this way, deme structure can lower the effective population size by lowerering the effective number of stochastic transmissions. For instance, in the example in Figure 3.3c, there are 20 individuals, but only 3 potential stochastic transmissions.

To check our intuition that deme structure can decrease the effective population size and increase genetic drift, we ran simulations of a simplified deme model (see Methods): all demes have the same number of individuals, and there is a sufficiently large enough number of demes that the total number of demes does not matter. Initially some number of demes are infected, and transmission occurs such that the overall effective reproduction number in the population is around 1. From our simulations, we find that when the number of individuals in a deme increases, the ratio between the number of infected individuals and the inferred scaled effective population size increases (Figure 3.3d); in other words, the more individuals there are in a deme, the higher the level of genetic drift we observe compared what is expected from the number of infected individuals. This is because while the number of infected individuals increases when the deme size increases (Figure 3.S12a), the inferred effective population size (and thus the level of stochasticity) stays the same as a function of deme size (it is more dependent on the number of infected demes) (Figure 3.S12b). However, the exact ratio of the number of infected individuals to the inferred effective size depends on the parameters of the model.

In reality, both superspreading and host structure are likely at play. Additionally, they could interact with each other. For instance, there could be superspreading within a deme. While we currently do not have sufficient data to tease apart the contribution of these two mechanisms, this would be an interesting and important avenue for future work. In particu-

lar, the relative contributions of the two mechanisms to genetic drift can affect evolutionary dynamics. Whereas increased genetic drift due to superspreading decreases establishment probability for new mutations because it is more likely for a new variant to go extinct, increased genetic drift due to host structure increases the establishment probability for new mutations because it is more likely for a well-connected individual to transmit to another well-connected individual [3].

3.4 Discussion

Here, we systematic studied the strength of genetic drift of SARS-CoV-2 in England across time and spatial scales. To do this, we developed and validated a method for jointly inferring genetic drift and overdispersed measurement noise using lineage frequency time series data, allowing these two effects to be disentangled, which overcomes a major challenge in the ability to infer the strength of genetic drift from time-series data. We find that the effective population size of SARS-CoV-2 in England was lower than that of a SEIR null model true population size (using the observed number of positives) by a time-dependent factor of 100 to 2350, suggesting that there were consistently high levels of genetic drift over time. We also find evidence for spatial structure in the transmission dynamics during the Alpha and Delta waves, as the inferred $\tilde{N}_e(t)$ was substantially lower in regions compared to that of England.

The levels of genetic drift that we observe are higher than literature values of superspreading, suggesting that additional mechanisms may be leading to increased stochasticity. In particular, we explore a simplified deme model with groups of individuals that are well-connected to one another (demes), and find that such a simple model can generate a low effective population size even in the absence of superspreading, due to jackpot events. Our results suggest that understanding heterogeneity in host population network structure may be important for understanding the high levels of genetic drift observed for SARS-CoV-2 in England.

Accurately estimating the strength of genetic drift allows us to better understand disease spread and extinction, as well as to better parameterize evolutionary models and understand how mutations will establish in the population. The amount by which genetic drift was elevated compared to the number of positives did not change much over time or across variants, despite changes in lockdowns and restrictions. On the other hand, this may not be so surprising given the findings that restrictions affect the mobility network structure in a complex way, decreasing some types of mobility while increasing others [18]. One exception was that Alpha had significantly higher genetic drift when compared to the number of positive Alpha cases when compared to Delta. This may be either due to differences in the properties of the virus or differences in host behavior.

We observe that measurement noise of SARS-CoV-2 tends to be more variable than uniform sampling when a variant first emerges, but is otherwise indistinguishable from uniform sampling. Our results suggest that joint inference of measurement noise and genetic drift

may be most important during the time when a variant first emerges. The number of SARS-CoV-2 sequences from England is extremely high and sampling biases are expected to be low, because of efforts to reduce sampling biases by sampling somewhat uniformly from the population through the COVID-19 Infection Survey [62] (from which a subset of positives are sequenced). On the other hand, other countries may have higher sampling biases, so jointly estimating measurement noise and genetic drift may be even more useful in analyzing data from these countries. It may also be interesting to analyze time series genomics data taken from wastewater, which we would expect to exhibit low measurement noise.

Limitations of the study

First, while we accounted for fitness differences between variants (i.e. pre-B.1.177, B.1.177, Alpha, and Delta) by analyzing them separately, we assumed that the lineages (either with the pango nomenclature or created by cutting the tree) within each of these groups of lineages or variants were neutral with respect to one another. This assumption is consistent with the current literature, but there could be deterministic changes in frequency due to small fitness differences between lineages or human behavior that are too weak to be detected with significance given the available data. Similarly, in our model we assumed that there was no mutation and migration. Introducing mutations with a small rate in the simulations did not have a large effect on the method performance, and we will explore the effect of migration in subsequent work. More generally, future work should explore joint inference of selection, migration, and/or mutation in the model, as appropriate for the pathogen of interest, building on previous work in this area [20, 39, 57].

Second, there may be biases in the way that data are collected that are not captured in our model. While our method does account for sampling biases that are uncorrelated in time, sampling biases that remain over time cannot be distinguished from genetic drift (i.e. if one geographical region was dominated by a particular lineage and it consistently had higher sequencing rates compared to another geographical region), and this can potentially bias the inferred effective population size; although, this is also a problem in phylogenetic methods. Additionally, we assume that the measurement noise overdispersion is identical for all lineages within a variant; in reality, there may be differences in sampling between lineages. Future work should explore the whether this is the case in the actual dataset, as well as the effect of lineage-specific measurement noise overdispersion on overall method performance.

Third, the quantity of effective population size is a summary statistic that is influenced by many factors, making its interpretation challenging. The effective population size describes the population size under a well-mixed Wright-Fisher model, whereas in reality, this assumption is broken by many effects, including host structure and broad offspring number distributions. Thus, in our study we are careful to interpret effective population size only in the broadest terms of genetic drift, without being able to determine what mechanisms lead to the inferred effective population size (although we do explore some possibilities above). While within-host dynamics may in principle impact the lineage frequency trajectories, this effect is likely small for our analysis because we focus on acute infections (infections in the

community rather than in hospitals and nursing homes). This is because acute infections of SARS-CoV-2 are thought to generate little within-host diversity that is passed on due to the short infection duration and small bottleneck size between hosts [37, 38]; while new mutations arising within acute hosts have been observed to be transmitted, these events are rare [37].

Comparison of time-series frequency method with common phylogenetics approaches

Currently, the primary approach used for inferring pathogen effective population size is phylogenetics. Phylogenetics methods arrange genomics sequences into a tree based on genomic distance and either measure the distribution of lineage sizes (number of sequences in different parts of the tree) [17] or fit the rate at which branches in the tree coalescence to determine the effective population size [56, 47, 64, 28].

The two approaches of inferring effective population size through time series frequency trajectories or phylogenetics have their respective benefits and drawbacks and can be chosen based on the particular application. A big benefit of the time-series frequency methods is the ability to detect and correct for time-varying measurement noise, which we have explored in this study. However, future work should explore how measurement noise affects phylogenetic trees. Time-series frequency methods also scale better computationally with the number of sequences, which is important for SARS-CoV-2 where there are have been millions of sequences collected and deposited into the GISAID repository. Another benefit of time-series frequency methods is that they only need a signal of standing variation, and can be used even when little genetic diversity accumulates, for instance for a species with a low mutation rate or in the recent past. Finally, time series inference methods are generally better able to incorporate different population genetics models, for instance models that include selection or broad offspring number distributions [43], although we do not utilize this benefit in our present analysis. The main drawback of time series frequency methods, including the method developed in this study, is that a sufficiently large number of sequences at high time resolution is needed to get a reasonable estimate of how the frequency changes over time. Additionally, the temporal resolution of the inferred parameters is limited; this may be improved by imposing a prior on how the parameters are expected to change over time.

Phylogenetics methods are particularly useful in situations with sparse data (they do not rely on frequencies) and high enough mutation rates such that sufficient genetic diversity is generated. However, the main drawbacks of phylogenetics methods are not being able to take into account complex measurement noise, being challenging to scale computationally for large numbers of sequences, being noisy when there is little genetic diversity, and assuming particular models of lineage coalescence (the Kingman coalescent is commonly assumed) which may not be a good assumption in some situations.

Time-series frequency methods and phylogenetics methods can also be used together.

If phylogenetic analyses are desired but measurement noise is an issue, then the method developed in this study can be used to detect and reject time periods with high measurement noise, and phylogenetics methods can subsequently be used on the filtered data. It would be valuable for future theoretical work to explore how to connect time series and phylogenetic methods, for instance the measurement noise parameter we fit here to the sampling parameter in phylogenetic birth-death models [56].

While we have focused on SARS-CoV-2 in this study, the method developed here can be extended to study genetic drift in other natural populations that are influenced by measurement noise and where genomic frequency data are available, for instance other pathogens, field studies, and ancient DNA [26, 15, 51]. More generally, ongoing methods development that integrates genomics, survey, and other data sources is crucial for being able to harness the large amounts of data that have been generated to better understand and predict evolutionary dynamics.

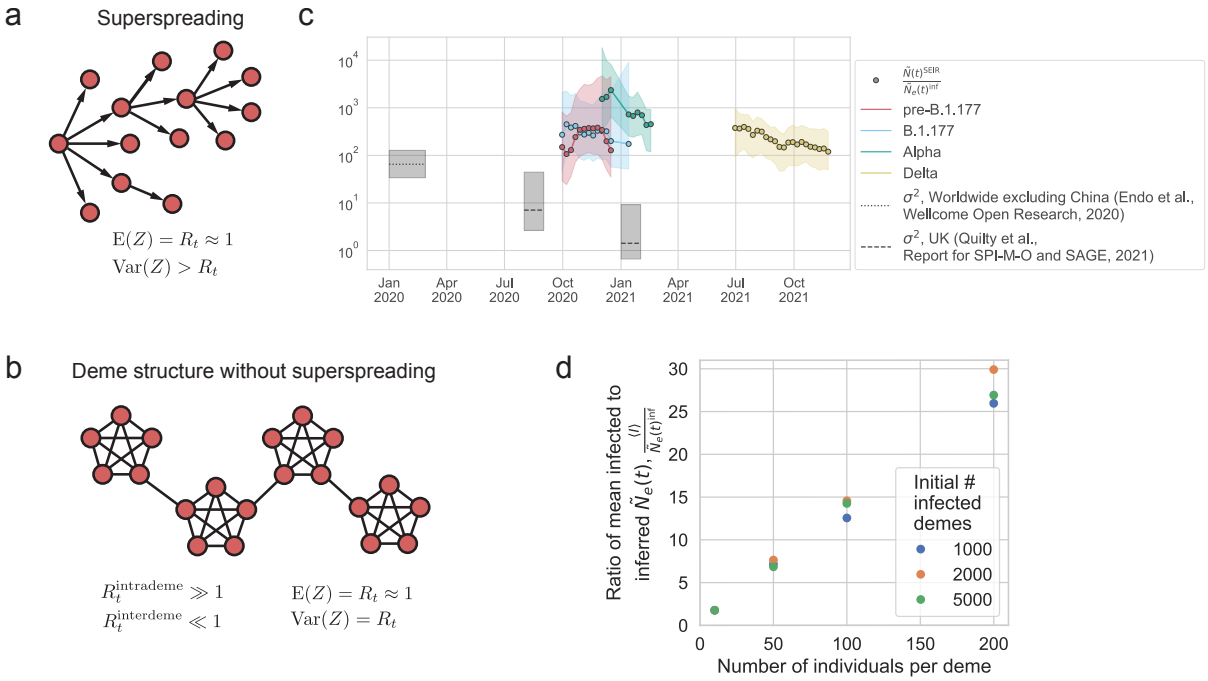

Figure 3.3: Potential mechanisms that can generate a low effective population size. (a) Superspreading, where the distribution of the number of secondary cases (Z) from a single infected individual is broadly distributed. (b) Deme structure without superspreading, due to heterogeneity in the host network structure. (c) The ratio between the $\tilde{N}^{\text{SEIR}}(t)$ (the scaled population size calculated from an SEIR model using the number of observed positive individuals and the observed effective reproduction number) and the inferred $\tilde{N}_e(t)$ for each variant. Only data where the error in the SEIR model $\tilde{N}^{\text{SEIR}}(t)$ is less than 3 times the value are shown, because larger error bars make it challenging to interpret the results. The infered $\tilde{N}_e(t)$ is lower than the $\tilde{N}^{\text{SEIR}}(t)$ (which assumes well-mixed dynamics and no superspreading) by a factor of 100 to 2350, indicating high levels of genetic drift. The variance in offspring number from the literature [19, 48] does not entirely explain the discrepancy between the true and effective population sizes. (d) Simulations of deme structure without superspreading can generate high levels of genetic drift via jackpot events. SEIR dynamics are simulated within demes (with $R_t = 10$, i.e. deterministic transmission) and Poisson transmission is simulated between demes ($R_t \ll 1$, i.e. stochastic transmission) such that the population $R_t \sim 1$ (see Methods). Simulation parameters are: mean transition rate from exposed to infected γ_E = (2.5 days)$^{-1}$, mean transition rate from infected to recovered γ_I = (6.5 days)$^{-1}$, total number of demes $D_{total} = 5.6 \times 10^5$. The ratio between the number of infected individuals and the inferred scaled effective population size is found to scale linearly with the deme size and not with the number of infected demes. This scaling results because of jackpot events where a lineage that happens to infect a susceptible deme grows rapidly until all susceptible individuals in the deme are infected.

3.5 Methods

Data sources and processing

We downloaded sequence data from the COVID-19 Genomics UK Consortium (COG-UK) [60]. This dataset is composed of a random sample of the positive cases from the COVID-19 Infection Survey, which is a surveillance study of positive individuals in the community administered by the Office for National Statistics (see below). For lineages that appeared before B.1.177, we downloaded the metadata from the COG-UK Microreact dashboard [12], which included the time and location of sample collection (UTLA), as well as the lineage designation (Pango nomenclature) [45, 50]. For B.1.177, Alpha, and Delta sequences, because the Pango nomenclature had very few lineages, we created our own lineages from the phylogenetic trees (see below). We downloaded the publicly available COG-UK tree on February 22, 2021 for B.1.177; June 20, 2021 for Alpha; and January 25, 2022 for Delta. We also downloaded the COG-UK metadata for all lineages on January 16, 2022, which included the time and location (UTLA) of sample collection. For the data of B.1.177, Alpha, and Delta, the data was deduplicated to remove reinfections in the same individual by the same lineage, but reinfections in the same individual by a different lineage were allowed.

The lineage frequency time-series is calculated separately for each variant or group of lineages (pre-B.1.177, B.1.177, Alpha, and Delta). First, the sequence metadata are aggregated by epidemiological week (Epiweek) to average out measurement noise that may arise due to variations in reporting within a week. Then, the lineage frequency is calculated by dividing the number of sequences from that lineage in the respective tree by the total number of sequences of that variant (or group of lineages) that were assigned to any lineage in the respective tree.

Because our model describes birth-death processes when the central limit theorem can be applied, we need the lineage frequencies to be sufficiently high. Thus, we randomly combine rare lineages into "superlineages" that are above a threshold number of counts and threshold frequency in the first and last timepoint of each trajectory. For the threshold, we chose of 20 counts and frequency of 0.01 because the inferred effective population size was roughly constant as a function of the threshold above this threshold value. Superlineages are non-overlapping (i.e. each sequence belongs to exactly one superlineage).

The estimated number of people testing positive for COVID-19 in England and each region of England was downloaded from the UK Office for National Statistics' COVID-19 Infection Survey [62]. The COVID-19 Infection Survey includes households that are semi-randomly chosen, and individuals are tested regardless of whether they are reporting symptoms. Infections reported in hospitals, care homes, and other communal establishments are excluded. Thus the dataset provides a representative number of positive individuals in the community setting. The reported date of positive cases is the date that the sample was taken. The error on the number of positive individuals from April 17, 2020 to July 5, 2020 is reported as the 95% confidence interval, and after July 5, 2020 is reported as the 95% credible interval. The regional data reported the positivity rate over two week intervals. To

get the number of positives, we multiplied by the number of individuals in the region. As the data was reported over two week intervals, we obtained the number of positives for each week using linear interpolation.

The observed effective reproduction numbers for England and each region of England were downloaded from the UK Health Security Agency [61]. Only times where the certainty criteria are met and the inference is not based on fewer days or lower quality data are kept. The error on the effective reproduction number is reported as the 90% confidence interval. Although not reported in the dataset, we choose the point estimate of the effective reproduction number to be the midpoint between the upper and lower bounds of the 90% confidence interval.

Creating lineages in B.1.177, Alpha, and Delta

For B.1.177, Alpha, and Delta, we divided each of them into neutral lineages based on phylogenetic distance. Specifically, for B.1.177 and Alpha, we cut a phylogenetic tree (in units of number of mutations) at a certain depths, $d = d_{\text{cut}}$. Each of the (internal or external) branches that are cut by the line $d = d_{\text{cut}}$ defines a lineage (Figure 3.2a). The (observed) frequency of a lineage at a given time point in England was computed by counting the number of England sequences (leaf nodes) belonging to the lineage and by normalizing it by the total number of sequences in all assigned lineages of the focal variant in England at that time point. Lineage frequencies at the regional level were similarly computed by counting the number of sequences separately for each region.

The choice of d_{cut} is arbitrary to some extent. Because we wanted a sufficiently high resolution of lineages from the early phase of spreading of a variant and because the evolutionary distance correlates with the actual sample date (Figure 3.S5), for each focal variant, we chose the depth d_{cut} that roughly corresponds to the time point when it began to spread over England.

For the Delta variant, the sequences form two distinct groups along the depth direction, as seen from the last panel of Figure 3.S5. Therefore, to divide the Delta variant into lineages with small frequencies, we cut the phylogenetic tree at two depths sequentially; we first cut the tree at $d_{\text{cut}}^{(1)}$, which resulted in lineages with small frequencies plus a lineage with $\mathcal{O}(1)$ frequency. Then, to divide the latter lineage further, we took the subtree associated with this lineage and cut the subtree at $d_{\text{cut}}^{(2)}$.

For the results presented in the main text, we used (in units of substitutions per site, with the reference d=0 being the most recent common ancestor) $d_{\text{cut}} = 2.323 \cdot 10^{-2}$ for B.1.177, $d_{\text{cut}} = 2.054 \cdot 10^{-3}$ for Alpha, and $d_{\text{cut}}^{(1)} = 1.687 \cdot 10^{-3}$ and $d_{\text{cut}}^{(2)} = 1.954 \cdot 10^{-3}$ for Delta. We confirmed that our results are robust to the choice of d_{cut} as well as the choice of the phylogenetic tree data we used (Figure 3.S7).

Model for inferring effective population size from lineage frequency time series

We use a Hidden Markov Model with continuous hidden and observed states to describe the processes of genetic drift and sampling of cases for sequencing (similar to a Kalman filter) (Figure 3.1A). The hidden states describe the true frequencies of the lineages and the observed states describe the observed frequencies of the lineages as measured via sequenced cases.

The transition probability between the true frequencies f_t (the hidden states) due to genetic drift when $0 \ll f \ll 1$ has been shown in [5] to be well-described by the following expression, which we use as our transition probability,

$$p(f_{t+1}|f_t, \tilde{N}_e(t)) = \frac{1}{2}\sqrt{\frac{2f_t^{1/2}}{\pi f_{t+1}^{3/2}(\tilde{N}_e(t))^{-1}}}\exp\left(-\frac{2(\sqrt{f_{t+1}}-\sqrt{f_t})^2}{(\tilde{N}_e(t))^{-1}}\right). \tag{3.4}$$

$\tilde{N}_e(T) \equiv N_e(t)\tau(t)$ where $N_e(t)$ is the time-dependent effective population size and $\tau(t)$ is the time-dependent generation time, which is defined as the mean time between two subsequent infections per individual (i.e. the time between when an individual becomes infected and infects another individual, or the time between two subsequent infections caused by the same individual). This transition probability gives the correct first and second moments describing genetic drift when $f \ll 1$, $\mathrm{E}(f_{t+1}|f_t) = f_t$ and $\mathrm{Var}(f_{t+1}|f_t) = \frac{f_t}{\tilde{N}_e(t)}$, and is a good approximation when the central limit theorem can be applied, which is the case when $f \gg 0$. By assuming that $f_{t+1} \approx f_t$, and defining $\phi_t \equiv \sqrt{f_t}$, Equation 3.4 can be approximated as a simple normal distribution

$$p(\phi_{t+1}|\phi_t, \tilde{N}_e(t)) = \mathcal{N}\left(\phi_t, \frac{1}{4\tilde{N}_e(t)}\right). \tag{3.5}$$

We describe the emission probability from the true frequency f_t to the observed frequency f_t^{obs} (the observed states), defining $\phi_t^{obs} \equiv \sqrt{x_t^{obs}}$, as

$$p(\phi_t^{obs}|\phi_t, c_t) = \mathcal{N}\left(\phi_t, \frac{c_t}{4M_t}\right) \tag{3.6}$$

where M_t is the number of input sequences. Again, this distribution is generically a good description when the number of counts is sufficiently large, due to the central limit theorem. The first and second moments of this emission probability are $\mathrm{E}(f_t^{obs}|f_t) = f_t$ and $\mathrm{Var}(f_t^{obs}|f_t) = \frac{c_t}{M_t}f_t$, or equivalently considering the number of sequences $n_t^{obs} = f_t^{obs}M_t$ and the true number of positive individuals n_t, $\mathrm{E}(n_t^{obs}|n_t) = n_t$ and $\mathrm{Var}(n_t^{obs}|n_t) = c_t n_t$. Thus, c_t describes the strength of measurement noise at time t. When $c_t = 1$, the emission probability approaches that describing uniform sampling of sequences from the population of positive individuals (i.e. can be described by a Poisson distribution in the limit of a large number of sequences), namely $\mathrm{Var}(n_t^{obs}|n_t) = n_t$ or equivalently $\mathrm{Var}(f_t^{obs}|f_t) = \frac{f_t}{M_t}$. This is the realistic

minimum amount of measurement noise. When $c_t > 1$, it describes a situation where there is bias (that is uncorrelated in time) in the way that sequences are chosen from the positive population. The case of $0 < c_t < 1$ describes underdispersed measurement noise, or noise that is less random than uniform sampling. The case of $c_t = 0$ describes no measurement noise (for instance, when all cases are sampled for sequencing). These last two situations is unlikely in our data, and thus as we describe below, we constrain $c_t \geq 1$ in the inference procedure. In addition to being a good description of measurement noise, defining the emission probability in the same normal distribution form as the transmission probability allows us to easily derive an analytical likelihood function, described below (Note: see Ref. [57] for a method to derive an analytical likelihood function for arbitrary forms of the transition and emission probabilities).

We derive the likelihood function (up to a constant) for the the Hidden Markov Model using the forward algorithm, although it can alternatively be derived by marginalizing over all hidden states. We assume an (improper) uniform prior on ϕ_0 (i.e. no information about the initial true frequency of the lineage).

$$p(\phi_0, \phi_0^{obs}, \theta_0) = p(\phi_0^{obs}|\phi_0, c_0)p(\phi_0) \tag{3.7}$$

$$p(\phi_0) \propto 1 \tag{3.8}$$

$$p(\phi_t, \phi_{0:t}^{obs}, \theta_{0:t}) = p(\phi_t^{obs}|\phi_t, c_t) \int_{-\infty}^{\infty} p(\phi_t|\phi_{t-1}, \tilde{N}_e(t))p(\phi_{t-1}, \phi_{0:t-1}^{obs}, \theta_{0:t-1})d\phi_{t-1}, \quad 0 < t \leq T \tag{3.9}$$

$$p(\phi_{0:T}^{obs}, \theta_{0:T}) = \int_{-\infty}^{\infty} p(\phi_T, \phi_{0:T}^{obs}, \theta_{0:T})d\phi_T \tag{3.10}$$

$$\mathcal{L}(\vec{\phi}_{0:T}^{obs}|\theta_{0:T}) = \prod_{\alpha} p(\{\phi_{0:T}^{obs}\}_\alpha, \theta_{0:T})p(\theta_{0:T}) \tag{3.11}$$

$$p(\theta_{0:T}) \propto 1 \tag{3.12}$$

$$\mathcal{L}(\vec{\phi}_{0:T}^{obs}|\theta_{0:T}) = \prod_{\alpha} p(\{\phi_{0:T}^{obs}\}_\alpha, \theta_{0:T}). \tag{3.13}$$

where $\phi_{0:t}^{obs} \equiv \{\phi_0^{obs}, ..., \phi_t^{obs}\}$, $\theta_{0:t} \equiv \{\tilde{N}_e(0), ..., \tilde{N}_e(t), c_0, ..., c_t\}$, and the subscript α indicates a particular lineage. We use a uniform prior on the parameters. The parameters $\theta_{0:T}$ are inferred by maximizing the likelihood (described below).

The forward algorithm has an analytical form for the simple case of Gaussian transition and emission probabilities. We use the identity for the product of two normal distributions $N(x, \mu, v)$, where μ is the mean and v is the variance:

$$N(x, \mu_1, v_1)N(x, \mu_2, v_2) = N(\mu_1, \mu_2, v_1 + v_2)N(x, \mu_{12}, v_{12}) \tag{3.14}$$

$$\mu_{12}(\mu_1, \mu_2, v_1, v_2) = \frac{\mu_1 v_2 + \mu_2 v_1}{v_1 + v_2} \tag{3.15}$$

$$v_{12}(v_1, v_2) = \frac{1}{\frac{1}{v_1} + \frac{1}{v_2}}. \tag{3.16}$$

Solving the forward algorithm recursively, we have

$$p(\phi^{obs}_{0:T}, \theta_{0:T}) = \prod_{i=1}^{T} N(\phi^{obs}_i, \mu_i, \frac{c_i}{4M_i} + v_i) \tag{3.17}$$

$$\tag{3.18}$$

where

$$\mu_1 = \phi^{obs}_0 \tag{3.19}$$

$$v_1 = \frac{\frac{1}{\tilde{N}_e(t)} + \frac{c_0}{M_0}}{4} \tag{3.20}$$

$$\mu_{i+1} = \mu_{12}(\mu_i, \phi^{obs}_i, v_i, \frac{c_i}{4M_i}) \tag{3.21}$$

$$v_{i+1} = v_{12}(\frac{c_i}{4M_i}, v_i) + \frac{1}{4\tilde{N}_e(t)}. \tag{3.22}$$

$$\tag{3.23}$$

Equation 3.17 can be substituted into Equation 3.13 to obtain the full analytical likelihood function.

Fitting the model to data

We split the time series data into overlapping periods of 9 Epiweeks, over which the effective population size is assumed to be constant. We first use the moments of the probability distributions combined with least squares minimization to get an initial guess for the parameters. Then, we perform maximum likelihood estimation using the full likelihood function. To capture uncertainties that arise from the formation of superlineages from lineages, we create superlineages randomly 100 times. We infer the strength of measurement noise and the effective population size for each superlineage combination (described below).

Determining the initial guess for the parameters using method of moments approach

Combining the transition and emission probabilities, and marginalizing over the hidden states we have

$$p(f^{obs}_j | f^{obs}_i) \propto \sqrt{\frac{1}{(f^{obs}_j)^{3/2}}} \exp\left(- \frac{2\left(\sqrt{f^{obs}_j} - \sqrt{f^{obs}_i}\right)^2}{\kappa_{i,j}} \right) \tag{3.24}$$

$$p(\phi^{obs}_j | \phi^{obs}_i) = \mathcal{N}(\phi^{obs}_i, \kappa_{i,j}) \tag{3.25}$$

$$\kappa_{i,j} \equiv \frac{c_i}{4M_i} + \frac{c_j}{4M_j} + \frac{(j-i)}{4\tilde{N}_e(t)}. \tag{3.26}$$

The first two terms of $\kappa_{i,j}$ are the contribution to the variance from measurement noise at times i ad j, and the third term is the contribution to the variance from genetic drift.

We calculate the maximum likelihood estimate of $\kappa_{i,j}$, $\hat{\kappa}_{i,j}$, which is simply the mean squared displacement

$$\hat{\kappa}_{i,j} = \left\langle (\phi_j^{obs} - \phi_i^{obs})^2 \right\rangle. \tag{3.27}$$

The standard error is given by

$$\Delta\hat{\kappa}_{i,j} = \sqrt{\frac{\left\langle \left[(\phi_j^{obs} - \phi_i^{obs})^2 - \hat{\kappa}_{i,j} \right]^2 \right\rangle}{Z}} \tag{3.28}$$

where Z is the number of superlineages.

By looking across all pairs of timepoints i and j, we get a system of linear equations in $\kappa_{i,j}$ that depend on the parameters c_t and $\tilde{N}_e(t)$. To determine the most likely values of the parameters, we minimize

$$\ln \sum_{i,j} \frac{(\hat{\kappa}_{i,j} - Ac)^2}{\Delta\hat{\kappa}_{i,j}} \tag{3.29}$$

using scipy.optimize.minimize with the L-BFGS-B method and the bounds $1 \leq c_t \leq 100$ and $1 \leq \tilde{N}_e(t) \leq 10^7$. While underdispersed measurement noise ($c_t < 1$) is in principle possible, we constrain $c_t \geq 1$ because realistically, the lowest amount of measurement noise will be from uniform sampling of sequences.

Maximum likelihood estimation of the parameters

For each set of superlineages, we use the inferred measurement noise values (c_t) and inferred scaled effective population size from above ($\tilde{N}_e(t)$) as initial guesses in the maximization the likelihood function in Equation 3.13 over the parameters. For the optimization, we use scipy.optimize.minimize_scalar with the Bounded method and the bounds $1 \leq c_t \leq 100$ and $1 \leq \tilde{N}_e(t) \leq 10^{11}$. The inferred $\tilde{N}_e(t)$ is reported as the $\tilde{N}_e(t)$ for the midpoint of the 9 Epiweek period. The reported errors on $\tilde{N}_e(t)$ are the 95% confidence intervals which are calculated by using the likelihood ratio to get a p-value [2, 53]. We replace the likelihood

with the profile likelihood, which has the nuisance parameters, in this case $c_{0:T}$, profiled out:

$$p > 0.05 \tag{3.30}$$

$$p = \int I\left[\frac{\mathcal{L}_{\tilde{N}_e}(\hat{c}_{0:T}|\vec{\phi}^{obs}_{0:T})}{\mathcal{L}_{\tilde{N}'_e}(\hat{c}_{0:T}|\vec{\phi}^{obs}_{0:T})} > 1\right] P_{\tilde{N}'_e}(\hat{c}_{0:T}|\vec{\phi}^{obs}_{0:T}) d\tilde{N}'_e \tag{3.31}$$

$$\hat{c}_{0:T} = \arg\max_{c_{0:T}} \mathcal{L}_{\tilde{N}_e}(c_{0:T}|\vec{\phi}^{obs}_{0:T}) \tag{3.32}$$

$$P_{\tilde{N}'_e}(\hat{c}_{0:T}|\vec{\phi}^{obs}_{0:T}) \propto \mathcal{L}_{\tilde{N}'_e}(\hat{c}_{0:T}|\vec{\phi}^{obs}_{0:T}) p(\tilde{N}_e) \tag{3.33}$$

$$p(\tilde{N}_e) \propto 1 \tag{3.34}$$

$$\tag{3.35}$$

where I is an indicator function that equals one when the argument is true and zero otherwise, $\mathcal{L}_{\tilde{N}_e}(\hat{c}_{0:T}|\vec{\phi}^{obs}_{0:T})$ is the profile likelihood with the nuisance parameters (in this case) $c_{0:T}$ profiled out, $P_{\tilde{N}'_e}(\hat{c}_{0:T}|\vec{\phi}^{obs}_{0:T})$ is the posterior where we have used a uniform prior. We also tried a Jeffreys prior which is used for variance parameters, but it gave similar results on simulated data because it looked relatively flat over the values of $\tilde{N}_e(t)$ of interest. As the Jeffreys prior was more computationally expensive than the uniform prior and the two priors gave similar results, we used the uniform prior for the analyses.

The reported values of c_t are the median across all superlineage combinations and across all time series segments where the timepoint appears. The reported errors on c_t are the 95% confidence intervals as calculated by the middle 95% of values across superlineage combinations and time series segments.

Correcting for the number of sequences assigned to lineages

Because some sequences occur before the cut point in the tree that is used for creating lineages, they are not included in any lineages. As a result, the number of sequences assigned to lineages is lower than the number of sequences in the tree. To correct for the bias in inferred effective population size that results from leaving out sequences from parts of the tree, we divide the inferred effective population size by the fraction of sequences in the tree that are assigned to a lineage. We note that while the number of sequences in the tree is less than the total number of sampled sequences, the sequences in the tree were chosen to be a representative fraction of the total sampled sequences. Thus, we do not need to additionally correct for the downsampling of sequences that were included in the tree. To test that randomly subsampling sequences for the analysis does not affect the results, we randomly subsampled half of the Delta sequences, and reran the analyses; the inferred effective population size was very similar to that from the full number of sequences (Figure 3.S8).

Simulations

For the model validation, we perform simulations of the lineage trajectories using a discrete Wright-Fisher model. 500 lineages are seeded initially, and the initial frequency of lineages is taken to be the same across all lineages. In each subsequent Epiweek, the true number of counts for a lineage is drawn from a multinomial distribution where the probabilities of different outcomes are the true frequencies of the lineages in the previous Epiweek and the number of experiments is the effective population size. The true frequency is calculated by dividing the true number of counts by N. The observed counts are drawn from a negative binomial distribution,

$$p(n_t^{obs}|f_t) = NB(r,q) \equiv \binom{n_t^{obs}+r-1}{r-1} q^r (1-q)^{n_t^{obs}} \tag{3.36}$$

$$r = \frac{f_t M_t}{c_t - 1} \tag{3.37}$$

$$q = \frac{1}{c_t} \tag{3.38}$$

which has the same mean and variance as the emission probability in Equation 3.6. The total number of observed sequences in each timepoint is calculated empirically after the simulation is completed, as it may not be exactly M_t. The simulation is run for 10 weeks of "burn-in" time before recording to allow for equilibration. Superlineages are created in the same way as described above.

For long time series simulations, some lineages will go extinct due to genetic drift, making it challenging to have sufficient data for the analysis. To be able to have a high enough number of lineages for the entire time series, we introduce mutations with a small rate $\mu = 0.01$ per generation per individual.

Calculating the effective population size for an SIR or SEIR model

The effective population size times the generation time in an SIR model is given by Refs. [30, 65]

$$\tilde{N}_e(t) \equiv N_e(t)\tau(t) = \frac{I(t)}{2R_t\gamma_I}. \tag{3.39}$$

For an SEIR model, we calculated $\tilde{N}_e(t)$ following the framework from Ref. [22]. This was done by considering how the mean number of lineages, A, changes going backwards in time, s, which is given by

$$\frac{dA}{ds} = -fp_c \tag{3.40}$$

where f is the number of transmissions per unit time and p_c is the probability that a transmission results in a coalescence being observed in our sample. p_c is given by the number

of ways of choosing two lineages divided by the number of ways of choosing two infectious individuals

$$p_c = \frac{\binom{A(s)}{2}}{\binom{N(s)}{2}} \underset{\lim_{N(s)\to\infty}}{=} \binom{A(s)}{2} \frac{2}{N(s)^2}. \tag{3.41}$$

where the limit assumes that the number of infectious individuals, $N(s)$, is large. In the Kingman coalescent we also have

$$\frac{dA}{ds} = -\binom{A(s)}{2} \frac{1}{\tilde{N}_e(t)}. \tag{3.42}$$

Combining Equations 3.40, 3.41, and 3.42, we have

$$\tilde{N}_e(t) = \frac{N(s)^2}{2f}. \tag{3.43}$$

Thus by determining the number of transmissions per unit time, f, and the number of infectious individuals, $N(s)$, in an SEIR model, we can find an expression for $\tilde{N}_e(t)$.

These quantities can be derived from the equations describing the number of susceptible (S), exposed (E), infectious (I), and recovered (R) individuals in an SEIR model

$$\frac{dS}{dt} = -\beta I \frac{S}{N_H} \tag{3.44}$$

$$\frac{dE}{dt} = \frac{\beta I S}{N_H} - \gamma_E E - \delta_E E \tag{3.45}$$

$$\frac{dI}{dt} = \gamma_E E - \gamma_I I - \delta_I I \tag{3.46}$$

$$\frac{dR}{dt} = \gamma_I I \tag{3.47}$$

where β is the number of transmissions per infectious individual per unit time (the number of contacts made by an infectious individual per unit time multiplied by the probability that a contact results in a transmission), N_H is the total population size ($N_H = S + E + I + R$), γ_E is the rate that an exposed individual becomes infectious, δ_E is the rate of death for an exposed individual, γ_I is the rate than an infectious individual recovers, and δ_I is the rate of death for an infectious individual.

The number of infectious individuals in a generation, $N(s)$, is given by the instantaneous number of infectious individuals plus the number of exposed individuals that will become infectious in that generation [30]. Thus,

$$N(s) = \frac{\gamma_E}{\gamma_E + \delta_E} E + I. \tag{3.48}$$

The number of transmissions per unit time is given by

$$f = \beta I \frac{S}{N_H}. \tag{3.49}$$

We rewrite f in terms of the effective reproduction number (for which data are available) which is given by the number of transmissions per unit time (f) divided by the number of recoveries and deaths per unit time

$$R_t = \frac{f}{(\gamma_I + \delta_I)I + \delta_E E}. \tag{3.50}$$

Putting everything together, we have that $\tilde{N}_e(t)$ for an SEIR model is given by

$$\tilde{N}_e(t) = \frac{\left[\left(\frac{\gamma_E}{\gamma_E+\gamma_I}\right)E + I\right]^2}{2R_t[(\gamma_I + \delta_I)I + \delta_E E]}. \tag{3.51}$$

For SARS-CoV-2, the death rates are much lower than the rate at which exposed individuals become infectious and the rate at which infectious individuals recover ($\delta_E, \delta_I \ll \gamma_E, \gamma_I$). In this limit, Equation 3.51 simplifies to

$$\tilde{N}_e(t) = \frac{(E + I)^2}{2R_t\gamma_I I}. \tag{3.52}$$

We use the estimated number of positives from the COVID-19 Infection Survey for $I(t)$. This number is an estimate of the number of positive individuals in the community as measured by surveillance and includes both symptomatic and asymptomatic individuals. While the estimated number of positives does not include cases from hospitals, care homes, and other communal establishments, community cases likely contribute the most to transmission. We used the measured effective reproduction number from the UK Health Security Agency for R_t.

To calculate the number of exposed individuals for the SEIR model, we solved for E in Equation 3.46 (taking $\delta_E \ll \gamma_E$)

$$E = \frac{1}{\gamma_E}\left(\frac{dI}{dt} + \gamma_I I\right). \tag{3.53}$$

$\frac{dI}{dt}$ was calculated numerically as $\frac{I(t+\Delta t)-I(t-\Delta t)}{2\Delta t}$ where $\Delta t = 1$ week. The parameter values used were $\gamma_E^{-1} = 3$ days and $\gamma_I^{-1} = 5.5$ days [41, 27]. The error on E was calculated by taking the minimum and maximum possible values from the combined error intervals of $I(t + \Delta t)$ and $I(t - \Delta t)$ (note that this does not correspond to a specific confidence interval size).

The error on $\tilde{N}_e(t)$ for the SEIR model was calculated similarly by taking the minimum and maximum possible values from the combined error intervals of E, I, and R_t. Only time points where the error interval of $\tilde{N}_e(t)$ was less than 3 times the value were kept.

Calculating the effective population size for an SEIR model by variant

To calculate the effective population size for an SEIR model by variant, we needed to determine the variant-specific: number of infectious individuals $I(t)$, number of exposed individuals $E(t)$, effective reproduction number R_t, and rate than an infectious individual recovers

γ_I. We assumed that γ_I is constant between variants because we were unable to find literature studies suggesting otherwise. We calculated the number of infectious individuals $I(t)$ by multiplying the total number of positives by the fraction of each variant in the reported sequences. This should be a good representation of the fraction of the variant in the population as the sequences are a random sample of cases detected via surveillance, with the exception of December 2020 when number of sequenced Alpha cases was downsampled due to Spike gene target failure which was unknown at the time, and in January 2021 when the number of Alpha sequences was upsampled to try to make up for unknown downsampling in the previous month. We calculated the number of variant-specific exposed individuals $E(t)$ in the same way as described above using the variant-specific number of infectious individuals. We assumed that the rate an exposed individual becomes infectious γ_E is constant between variants.

We calculated the variant-specific effective reproduction number by rescaling the measured effective reproduction number for the whole population

$$R_t^v = R_t \frac{R_0^v}{\sum_w R_0^w f^w} \tag{3.54}$$

where R_0^w is the basic reproduction number of the variant w and f^w is the fraction of the infectious population with variant w. The values of R_0 are taken from Ref. [9], and when rescaled to $R_0^{pre-B.1.1.7}$ are $\frac{R_0^{pre-B.1.1.7}}{R_0^{pre-B.1.1.7}} = 1$, $\frac{R_0^{Alpha}}{R_0^{pre-B.1.1.7}} = 1.29$, $\frac{R_0^{Delta}}{R_0^{pre-B.1.1.7}} = 1.97$.

Deme simulations

To better understand the effect of host population structure on the effective population size, we simulated a simple situation where there are "demes", or groups, of individuals with very high rates of transmission between individuals in that deme, but the rate of transmission between individuals from different demes is very low. In a given simulation, all demes have the same number of individuals (10, 50, 100, or 200). The total number of demes is chosen to be very high (5.6×10^6). Initially, a certain number of demes (100, 1000, 2000, or 5000) are each seeded by a single infectious individual infected by a randomly chosen lineage (200 different lineages). We simulated deterministic SEIR dynamics within demes with $R_0 = 10$, $\gamma_E = (2.5 \text{ days})^{-1}$, $\gamma_I = (6.5 \text{ days})^{-1}$. We simulated Poisson transmission dynamics between demes. In order to calibrate the overall population dynamics to be roughly in equilibrium (the number of infectious individuals is not deterministically growing or shrinking), we draw the number of between-deme infections caused by a given deme from a Poisson distribution with mean 1. The time of the between-deme infection event is randomly chosen, weighted by the number of infected individuals within a deme at a given time. The number of infectious individuals in each lineage is recorded every 1 week, and the frequency of the lineage is calculated by dividing by the total number of infectious individuals from all lineages in that week. The lineage frequency data from a period of 9 weeks starting in week 42 is used for the inference of effective population size. The effective population size inference is performed as above except in the absence of measurement noise, so there is no emission step.

Acknowledgements

We are grateful to the Hallatschek lab for helpful discussions, feedback, and comments on the manuscript, particularly Giulio Issachini and Valentin Slepukhin. We thank Mike Boots, Vince Buffalo, Katia Koelle, Priya Moorjani, Rasmus Nielsen, Daniel Reeves, and Daniel Weissman for helpful discussions and feedback. We thank Hernan G. Garcia and Yun S. Song for helpful comments on the manuscript. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award R01GM115851, the National Science Foundation CAREER Award (#1555330), and the Simons Investigator award from the Simons Foundation (#327934). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC BER-ERCAP0019907.

3.6 Supplementary information

Summary of existing methods for inferring the strength of genetic drift

There are currently four main types of methods for estimating the strength of genetic drift in pathogen transmission, which we summarize here for giving context to this study.

1. **Contact tracing** can directly measure superspreading by following the close contacts of infected individuals to measure the distribution of the number of secondary cases (the offspring number distribution) [35]. However, some secondary cases may be missed which can lead to measurement bias. Additionally, it is challenging to trace multiple generations of transmission, so we miss important information on host contact network structure.

2. Another type of method fits disease prevalence over time to **branching process models** [19]. These models assume a particular distribution for the offspring number distribution (often a negative binomial distribution) and estimate the combination of parameters of the offspring number distribution along with growth rate that best fit the observed disease prevalence. External information about the growth rate can be used to constrain the parameters of the offspring number distribution.

3. **Phylogenetics** methods arrange genomics sequences into a tree based on genomic distance and either measure the distribution of lineage sizes (number of sequences in different parts of the tree) [17] or fit the rate at which branches in the tree coalescence to determine the effective population size [56, 47, 64, 28]. The effective population size is the population size that would reproduce the observed population dynamics under the idealized conditions of Wright-Fisher dynamics (discrete non- overlapping generations, a constant population size, and offspring determined by sampling with replacement from the previous generation). A lower effective population size indicates a higher level of genetic drift.

4. **Time series frequency methods** make use of a signature that genetic drift leaves in time series data, which is that it causes fluctuations in the lineage abundances. Higher amounts of genetic drift (lower effective population size) lead to larger fluctuations, and the magnitude of the fluctuations can be fit to determine the effective population size [67] (Figure 3.1a). Time series methods have also been used extensively in population genetics [7, 21, 20, 8, 57, 66] and to estimate within-host effective population size [36] and between-host transmission bottleneck sizes [55].

Application to COG-UK data by regions in England

The inference of effective population size can also reveal information about the well-mixed or spatially-structured nature of transmission dynamics within England. This can be done

Figure 3.S1: Inferred effective population size in regions of England. (Top panels) Inferred $\tilde{N}_e(t)$ of pre-B.1.177 lineages, B.1.177, Alpha, and Delta for each region of England. The inferred $\tilde{N}_e(t)$ for England as a whole is shown for reference. (Bottom panels) The ratio between the inferred $\tilde{N}_e(t)$ of England and that of the region for each variant. A horizontal dashed line indicates a ratio of 1 (i.e. $\tilde{N}_e(t)$ is the same in that region of England and England as a whole). Shaded regions show 95% confidence intervals (see Methods).

by inferring effective population size at smaller geographical scales within England. If the transmission dynamics were completely well-mixed, then we would expect $\tilde{N}_e(t)$ to be the same across regions and compared to England. On the other hand, if the transmission dynamics were completely spatially segregated (i.e. transmission only occurs within the defined geographical areas, but not between them) and the dynamics were the same in each

region, we would expect that the ratio $\tilde{N_e}^{\text{SEIR}}(t)/\tilde{N_e}^{\text{inf}}(t)$ to be the same across regions.

The geographical areas that we used were the 9 regions of England: East Midlands, East of England, London, North East, North West, South East, South West, West Midlands, and Yorkshire and The Humber. We looked at sequences from each region, repeating the analysis described above, and inferred the scaled effective population size (Figure 3.S1). In all regions, we observe a lower $\tilde{N}_e(t)$ for Delta in the region than in England, by a factor of up to 15. In some regions, the $\tilde{N}_e(t)$ for Alpha is lower in the regions than in England. For B.1.177 and and lineages pre-B.1.177, in some regions and timepoints the inferred $\tilde{N}_e(t)$ is lower in the region than in England, but for most timepoints the inferred values are very noisy due to there being fewer sequences at the regional level. These results suggest that the dynamics are not well-mixed during the Alpha and Delta waves.

The calculated SEIR model $\tilde{N_e}^{\text{SEIR}}(t)$ was too noisy to be used for comparison with the inferred $\tilde{N}_e(t)$. Instead, we compared the inferred $\tilde{N}_e(t)$ to the number of positive individuals in each region which is similar to the SEIR $\tilde{N_e}^{\text{SEIR}}(t)$ when $R_t \approx 1$ as it is in England (Figure 3.S10). In most cases the number of positives was 1-2 orders of magnitude higher than the inferred $\tilde{N}_e(t)$, suggesting high levels of genetic drift. The ratios of the number of positives to the inferred $\tilde{N}_e(t)$ in the regions were similar to one another and to that seen in England as a whole, suggesting that the dynamics are spatially-structured.

Similarly to in England as a whole, the inferred measurement noise in each region for Alpha and Delta tended to be high when the variant first emerged (Figure 3.S11), but unlike for England as a whole, the measurement noise in each region for pre-B.1.177 lineages was low when the lineage first emerged. The exceptions where London and North West where the measurement noise was indistinguishable from uniform sampling in general. This may be due to better sequencing efforts in these two regions, as they have a higher population size. For the other regions at other times, the measurement noise was in general indistinguishable from uniform sampling of sequences except from in a few timepoints. The inability to distinguish the strength of sampling noise in most time points from that of uniform sampling is partly due to the large error which results from there being fewer sequences at the regional level.

Additional supplementary figures

Table 3.S1: Overdispersion values from the literature ordered from highest to lowest. Any error intervals that are reported are taken from the reference (sometimes defined differently). The estimate taken from Ref. [48] assumes no self-isolation upon symptom onset and no testing; lifting these assumptions leads to similar or lower overdispersion.

Date	**Location**	**Method**	$\langle Z \rangle$	**Var(Z)**	**Ref.**
Beginning of pandemic to February 27 2020	Worldwide excluding China	Branching process model of number of imported and local cases	2.5 (,)	65 (33.75, 127.5)	[19]
March 1 to May 3 2020	Georgia (USA)	Spatiotemporal transmission model fit to multiple data sources	2 (0.5, 3.5)	12.26 (0.88, 101.5)	[33]
March 1 to November 1 2020	Denmark	Model fitting the case numbers across multiple regions	1.1 (0.8, 1.4)	12.1 (4.36, 25.9)	[29]
Beginning of pandemic until January 18 2020	China (Wuhan)	Stochastic simulations fit to infected cases	2.2 (1.4, 3.8)	11.16 (1.68, 1035.2)	[52]
August to September 2020	UK	Model using empirical viral load trajectories and contact numbers	1.21 (0.84, 2.51)	7.07 (2.65, 44.51)	[48]
May 15 to August 1 2020	Tamil Nadu and Andhra Pradesh (India)	Contact tracing and incidence	1.25 (1.1, 1.4)	4.31 (3.43, 5.4)	[34]

January to February 2021	UK	Model using empirical viral load trajectories and contact numbers	0.54 (0.4, 1.03)	1.42 (0.66, 9.19)	[48]
January 23 to April 28 2020	Hong Kong	Contact tracing	0.58 (,)	1.36 (,)	[1]
January 16 to April 3 2020	Hunan (China)	Contact tracing	0.4 (0.35, 0.47)	0.93 (0.66, 1.43)	[58]
January 14 to February 12 2020	Shenzhen (China)	Contact tracing	0.4 (0.3, 0.5)	0.68 (0.38, 1.21)	[6]

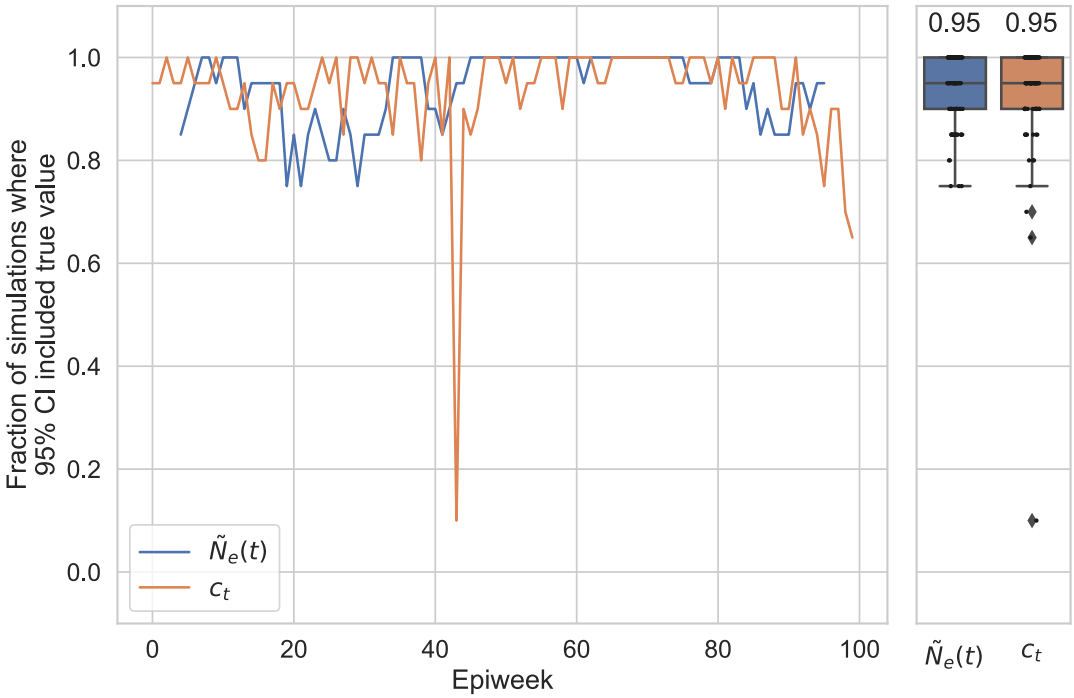

Figure 3.S2: The fraction of simulations (20 total) where the inferred 95% confidence interval for $\tilde{N}_e(t)$ or c included the true value (left) by timepoint and (right) for all timepoints. (Right) Boxes indicate the quartiles and the line inside the box (and number above) indicates the median. Whiskers indicate the extreme values excluding outliers. Simulation parameters are specified in the Methods and Figure 3.1.

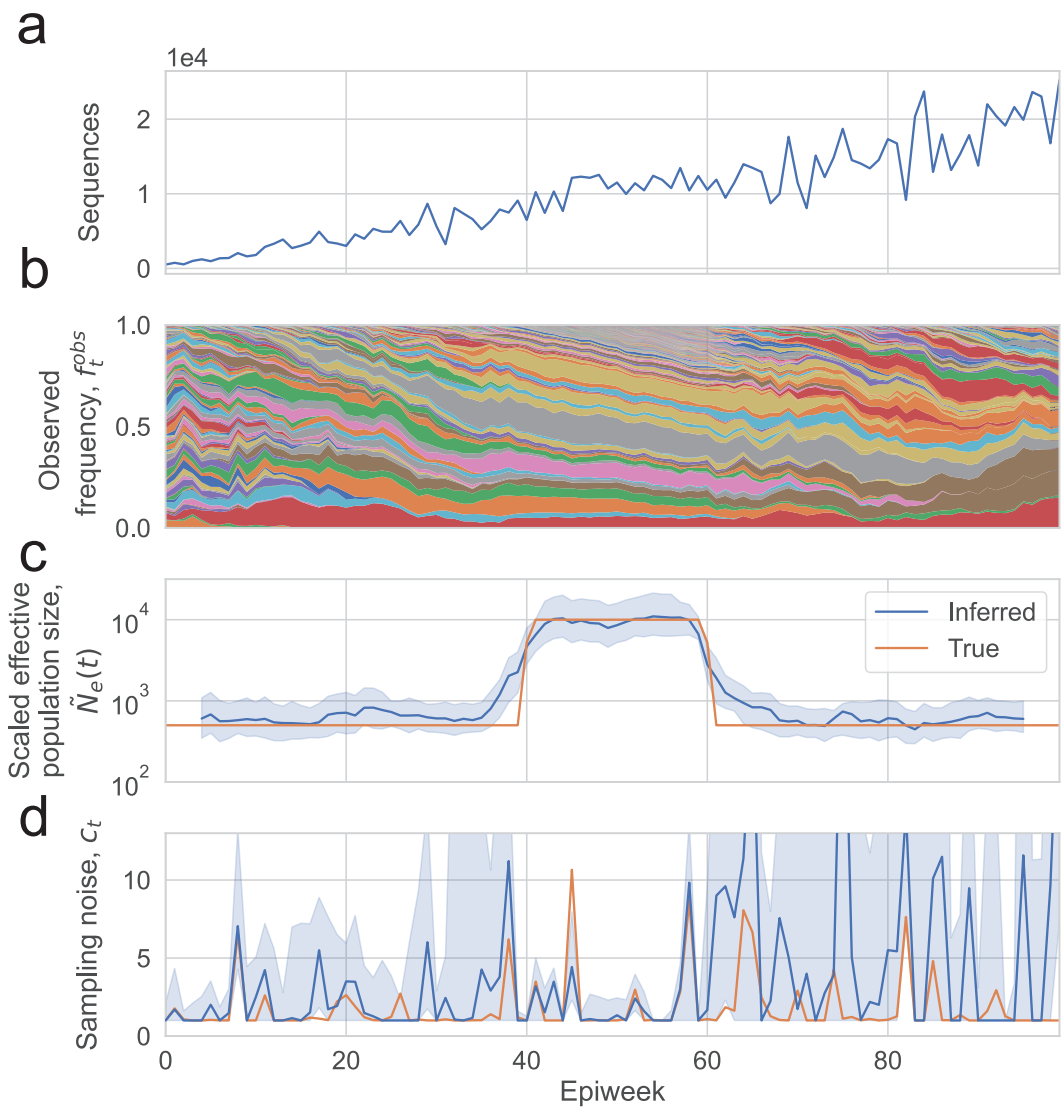

Figure 3.S3: Wright-Fisher simulations where $\tilde{N}_e(t)$ changes over time according to a rectangular function, and the inferred $\tilde{N}_e(t)$ and c_t. (a) Number of sequences sampled. (b) Simulated lineage frequency trajectories. (c) Inferred effective population size ($\tilde{N}_e(t)$) on simulated data compared to true values. (d) Inferred measurement noise (c_t) on simulated data compared to true values. In (c) the shaded region shows the 95% confidence interval calculated using the posterior, and in (d) the shaded region shows the 95% confidence interval calculated using bootstrapping (see Methods).

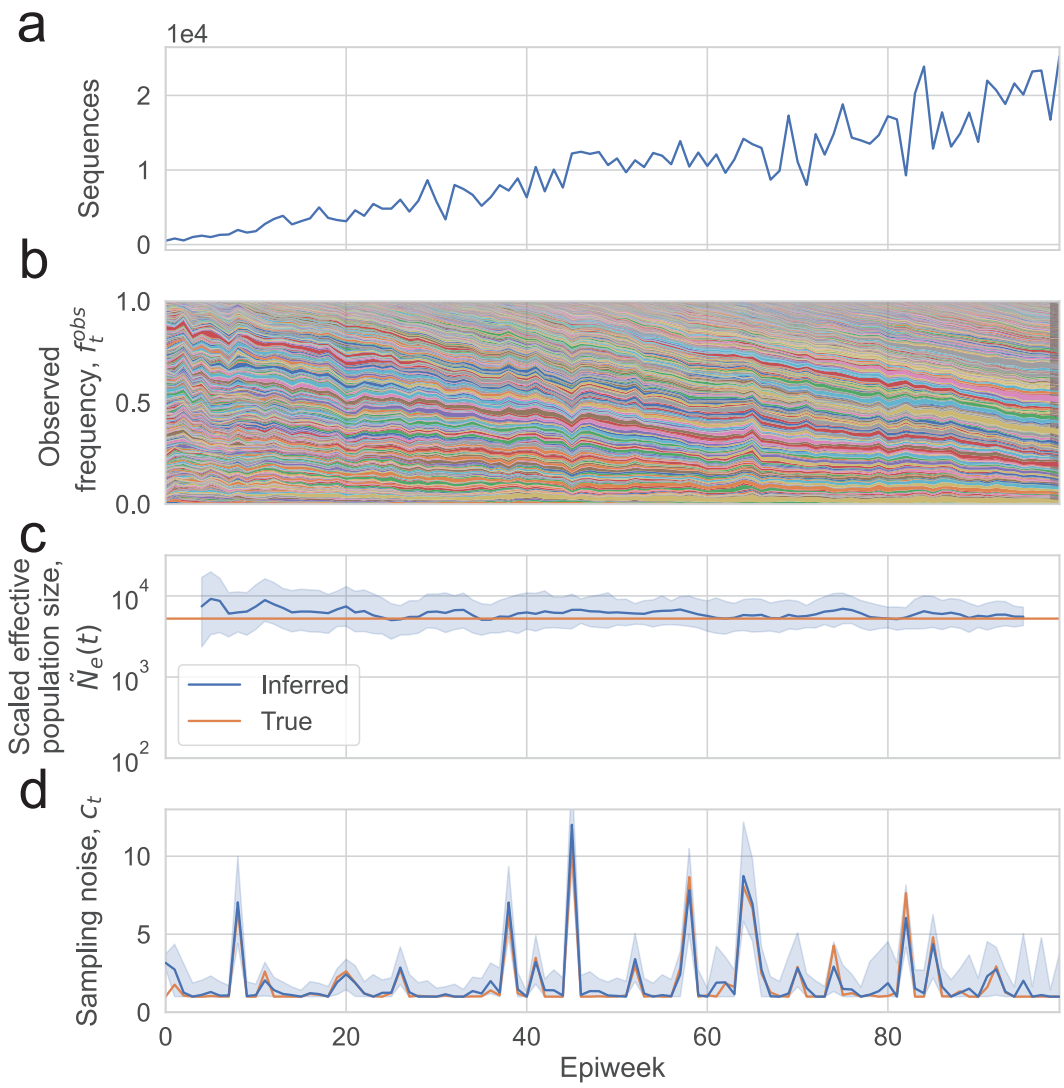

Figure 3.S4: Wright-Fisher simulations where $\tilde{N}_e(t)$ is constant over time, and the inferred $\tilde{N}_e(t)$ and c_t. (a) Number of sequences sampled. (b) Simulated lineage frequency trajectories. (c) Inferred effective population size ($\tilde{N}_e(t)$) on simulated data compared to true values. (d) Inferred measurement noise (c_t) on simulated data compared to true values. In (c) the shaded region shows the 95% confidence interval calculated using the posterior, and in (d) the shaded region shows the 95% confidence interval calculated using bootstrapping (see Methods).

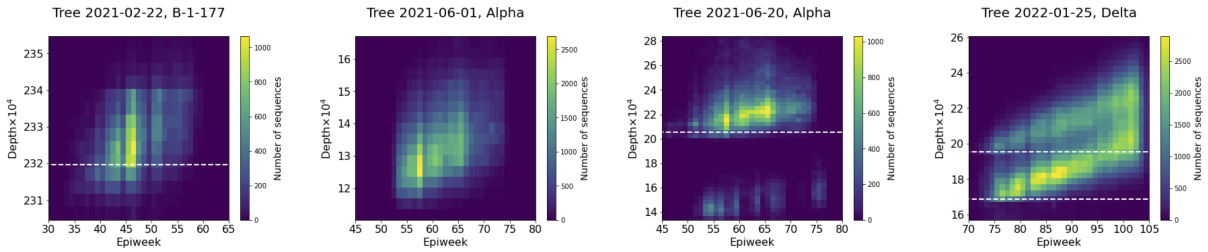

Figure 3.S5: Sample epiweeks versus tree depths. In a phylogenetic tree, the number of sequences (leaf nodes) of a focal variant that fall within specific epiweek and tree depth ranges is counted and summarized as a two-dimensional histogram. The tree depth is the substitution rate measured in units of substitutions per site, with respect to the most recent common ancestor. From left to right, the phylogenetic tree (specified by date) and focal variant are {2021-02-22, B-1-177}, {2021-06-01, Alpha}, {2021-06-20, Alpha}, and {2022-01-25, Delta}. Weeks are counted from 2019-12-29. The dashed horizontal lines indicate the values of d_{cut} ($d_{\text{cut}}^{(1)}$ and $d_{\text{cut}}^{(2)}$ for the Delta variant) used for the results presented in the main text.

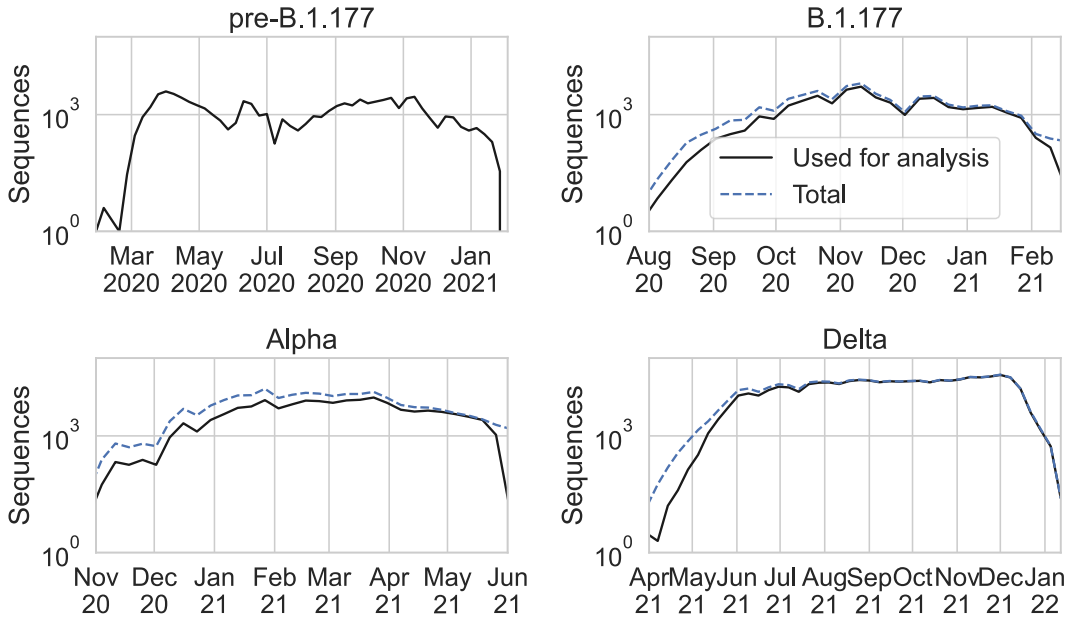

Figure 3.S6: Total number of sequences of each variant in the metadata from COG-UK downloaded on January 16, 2022 and the number of sequences used in the analysis for each variant or group of lineages (determined by the number of sequences included in the tree, and the number of sequences which could be grouped into sublineages based on the procedure described in the Methods).

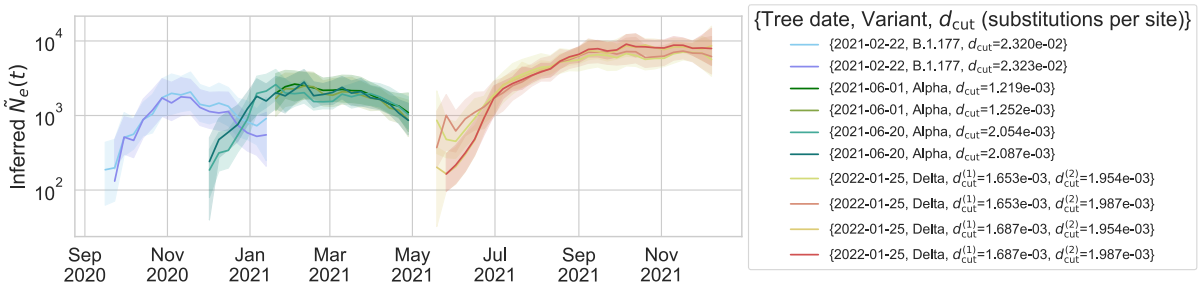

Figure 3.S7: Varying the date of the tree downloaded from COG-UK and the depth at which the tree is cut for creating lineages (d_{cut}, see Methods) does not substantially change the inferred scaled effective population size. The the tree date and depth used in the main text are $\{$2021-02-22, B.1.177, $d_{\text{cut}} = 2.323 \cdot 10^{-2}\}$, $\{$2021-06-20, Alpha, $d_{\text{cut}} = 2.054 \cdot 10^{-3}\}$, $\{$2022-01-25, Delta, $d_{\text{cut}}^{(1)} = 1.687 \cdot 10^{-3}$, $d_{\text{cut}}^{(2)} = 1.954 \cdot 10^{-3}\}$. The color of the lines for the parameters that were used in the main text are the same as those shown in Figure 3.2.

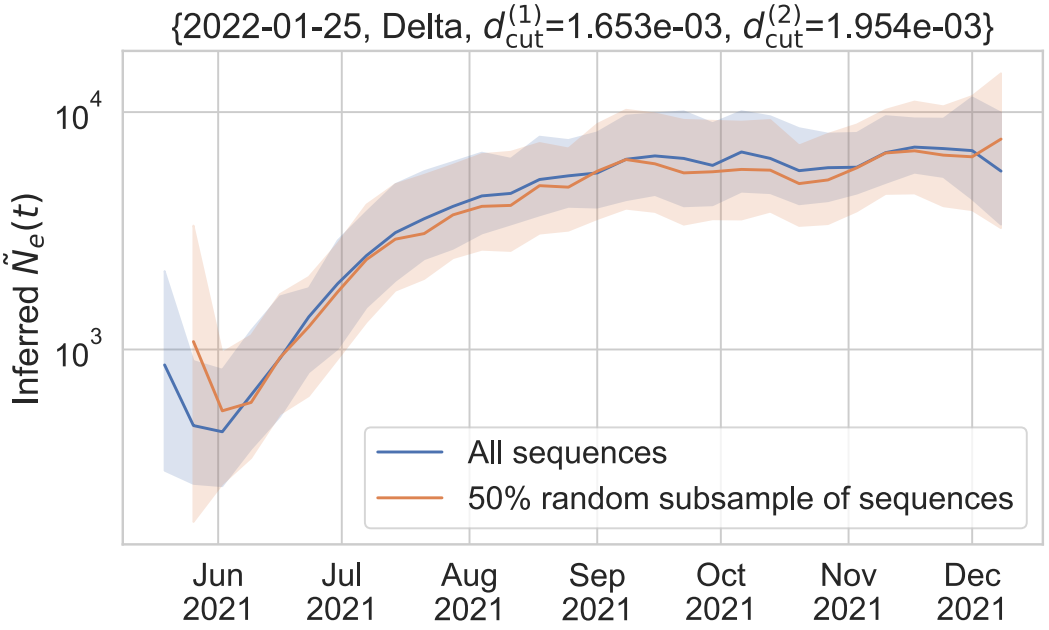

Figure 3.S8: Randomly subsampling half of the Delta sequences used for the analysis does not substantially change the inferred scaled effective population size.

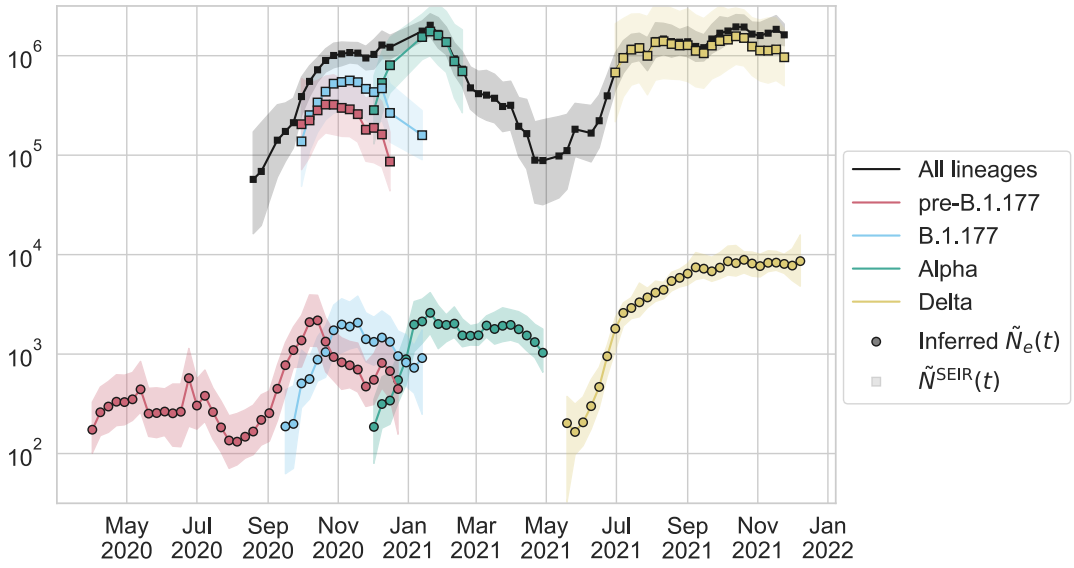

Figure 3.S9: Inferred scaled effective population size compared to the SEIR model scaled population size calculated using the observed number of positive individuals in England (see Methods).

Figure 3.S10: Inferred scaled effective population size by region in England, compared to number of positives at the community level in that region reported by the COVID-19 Infection Survey [62].

Figure 3.S11: Inferred measurement noise by region in England.

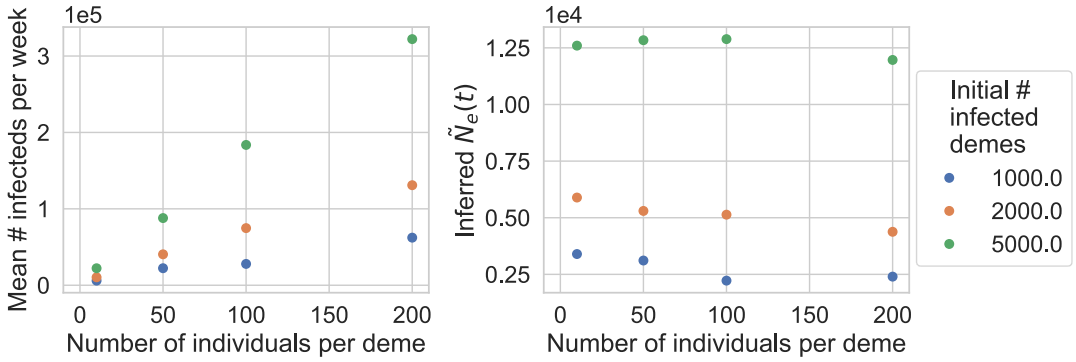

Figure 3.S12: Simulations of deme structure (described in main text and Methods). (a) The mean number of infected individuals per week from Weeks 42 to 50. (b) The inferred $\tilde{N}_e(t)$ using lineage trajectories from Weeks 42 to 50.

3.7 Bibliography

[1] D. C. Adam, P. Wu, J. Y. Wong, E. H. Lau, T. K. Tsang, S. Cauchemez, G. M. Leung, and B. J. Cowling. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine*, 26(11):1714–1719, 2020.

[2] M. Aitkin, R. J. Boys, and T. Chadwick. Bayesian point null hypothesis testing via the posterior likelihood ratio. *Statistics and Computing*, 15(3):217–230, 2005.

[3] H. Alexander and T. Day. Risk factors for the evolutionary emergence of pathogens. *Journal of The Royal Society Interface*, 7(51):1455–1474, 2010.

[4] B. M. Althouse, E. A. Wenger, J. C. Miller, S. V. Scarpino, A. Allard, L. Hébert-Dufresne, and H. Hu. Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. *PLoS Biology*, 18(11):e3000897, 2020.

[5] A. N. N. Ba, I. Cvijović, J. I. R. Echenique, K. R. Lawrence, A. Rego-Costa, X. Liu, S. F. Levy, and M. M. Desai. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature*, 575(7783):494–499, 2019.

[6] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*, 20(8):911–919, 2020.

[7] J. P. Bollback, T. L. York, and R. Nielsen. Estimation of $2N_e s$ from temporal allele frequency data. *Genetics*, 179(1):497–502, 2008.

[8] V. Buffalo and G. Coop. The linked selection signature of rapid adaptation in temporal genomic data. *Genetics*, 213(3):1007–1045, 2019.

[9] F. Campbell, B. Archer, H. Laurenson-Schafer, Y. Jinnai, F. Konings, N. Batra, B. Pavlin, K. Vandemaele, M. D. Van Kerkhove, T. Jombart, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance*, 26(24):2100509, 2021.

[10] R. Challen, L. Dyson, C. E. Overton, L. M. Guzman-Rincon, E. M. Hill, H. B. Stage, E. Brooks-Pollock, L. Pellis, F. Scarabel, D. J. Pascall, et al. Early epidemiological signatures of novel SARS-CoV-2 variants: establishment of B.1.617.2 in England. *medRxiv*, 2021.

[11] B. Charlesworth. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3):195–205, 2009.

[12] COG-UK and Microreact. UK SARS-CoV-2. https://microreact.org/project/mxgBucHEZCZgsSWNrnQQRo-%20uk-sars-cov-2-2020-02-052021-04-20. Accessed: 2021-04-20.

[13] N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, A. J. Kucharski, J. D. Munday, C. A. Pearson, T. W. Russell, D. C. Tully, A. D. Washburne, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538): eabg3055, 2021.

[14] T. Day, S. Gandon, S. Lion, and S. P. Otto. On the evolutionary epidemiology of SARS-CoV-2. *Current Biology*, 30(15):R849–R857, 2020.

[15] M. Dehasque, M. C. Ávila-Arcos, D. Díez-del Molino, M. Fumagalli, K. Guschanski, E. D. Lorenzen, A.-S. Malaspinas, T. Marques-Bonet, M. D. Martin, G. G. Murray, et al. Inference of natural selection from ancient DNA. *Evolution Letters*, 4(2):94–108, 2020.

[16] C. H. v. Dorp, E. E. Goldberg, N. Hengartner, R. Ke, and E. O. Romero-Severson. Estimating the strength of selection for new SARS-CoV-2 variants. *Nature Communications*, 12(1):1–13, 2021.

[17] L. Du Plessis, J. T. McCrone, A. E. Zarebski, V. Hill, C. Ruis, B. Gutierrez, J. Raghwani, J. Ashworth, R. Colquhoun, T. R. Connor, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*, 371(6530):708–712, 2021.

[18] P. Edsberg Møllgaard, S. Lehmann, and L. Alessandretti. Understanding components of mobility during the COVID-19 pandemic. *Philosophical Transactions of the Royal Society A*, 380(2214):20210118, 2022.

[19] A. Endo et al. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, 5, 2020.

[20] A. F. Feder, S. Kryazhimskiy, and J. B. Plotkin. Identifying signatures of selection in genetic time series. *Genetics*, 196(2):509–522, 2014.

[21] A. Ferrer-Admetlla, C. Leuenberger, J. D. Jensen, and D. Wegmann. An approximate Markov model for the Wright–Fisher diffusion and its application to time series data. *Genetics*, 203(2):831–846, 2016.

[22] S. D. Frost and E. M. Volz. Viral phylodynamics and the search for an 'effective number of infections'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1548):1879–1890, 2010.

[23] S. D. Frost, O. G. Pybus, J. R. Gog, C. Viboud, S. Bonhoeffer, and T. Bedford. Eight challenges in phylodynamic inference. *Epidemics*, 10:88–92, 2015.

[24] A. Goyal, D. B. Reeves, E. F. Cardozo-Ojeda, J. T. Schiffer, and B. T. Mayer. Viral load and contact heterogeneity predict SARS-CoV-2 transmission and super-spreading events. *eLife*, 10:e63537, 2021.

[25] A. Goyal, D. B. Reeves, and J. T. Schiffer. Early super-spreader events are a likely determinant of novel SARS-CoV-2 variant predominance. *medRxiv*, 2021.

[26] R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, et al. A draft sequence of the Neandertal genome. *Science*, 328(5979):710–722, 2010.

[27] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine*, 26(5):672–675, 2020.

[28] S. Y. Ho and B. Shapiro. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, 11(3):423–434, 2011.

[29] J. B. Kirkegaard and K. Sneppen. Variability of individual infectiousness derived from aggregate statistics of COVID-19. *medRxiv*, 2021.

[30] K. Koelle and D. A. Rasmussen. Rates of coalescence for common epidemiological models at equilibrium. *Journal of the Royal Society Interface*, 9(70):997–1007, 2012.

[31] A. Kucharski and C. L. Althaus. The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Eurosurveillance*, 20(25):21167, 2015.

[32] S. S. Lakdawala and V. D. Menachery. Catch me if you can: superspreading of COVID-19. *Trends in Microbiology*, 29(10):919–929, 2021.

[33] M. S. Lau, B. Grenfell, M. Thomas, M. Bryan, K. Nelson, and B. Lopman. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proceedings of the National Academy of Sciences*, 117(36):22430–22435, 2020.

[34] R. Laxminarayan, B. Wahl, S. R. Dudala, K. Gopal, C. Mohan B, S. Neelima, K. Jawahar Reddy, J. Radhakrishnan, and J. A. Lewnard. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science*, 370(6517):691–697, 2020.

[35] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[36] C. K. Lumby, L. Zhao, J. Breuer, and C. J. Illingworth. A large effective population size for established within-host influenza virus infection. *eLife*, 9:e56915, 2020.

[37] K. A. Lythgoe, M. Hall, L. Ferretti, M. de Cesare, G. MacIntyre-Cockett, A. Trebes, M. Andersson, N. Otecko, E. L. Wise, N. Moore, et al. SARS-CoV-2 within-host diversity and transmission. *Science*, 372(6539):eabg0821, 2021.

[38] M. A. Martin and K. Koelle. Comment on "Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2". *Science translational medicine*, 13(617):eabh1803, 2021.

[39] I. Mathieson and G. McVean. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics*, 193(3):973–984, 2013.

[40] Y. A. Melsew, M. Gambhir, A. C. Cheng, E. S. McBryde, J. T. Denholm, E. L. Tay, and J. M. Trauer. The role of super-spreading events in Mycobacterium tuberculosis transmission: evidence from contact tracing. *BMC Infectious Diseases*, 19(1):1–9, 2019.

[41] D. Miller, M. A. Martin, N. Harel, O. Tirosh, T. Kustin, M. Meir, N. Sorek, S. Gefen-Halevi, S. Amit, O. Vorontsov, et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *Nature Communications*, 11(1):1–10, 2020.

[42] T. Obadia, R. Silhol, L. Opatowski, L. Temime, J. Legrand, A. C. Thiébaut, J.-L. Herrmann, E. Fleury, D. Guillemot, P.-Y. Boelle, et al. Detailed contact data and the dissemination of Staphylococcus aureus in hospitals. *PLoS Computational Biology*, 11 (3):e1004170, 2015.

[43] T. Okada and O. Hallatschek. Dynamic sampling bias and overdispersion induced by skewed offspring distributions. *Genetics*, 219(4):iyab135, 2021.

[44] S. P. Otto, T. Day, J. Arino, C. Colijn, J. Dushoff, M. Li, S. Mechai, G. Van Domselaar, J. Wu, D. J. Earn, et al. The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Current Biology*, 31(14):R918–R929, 2021.

[45] Á. O'Toole, E. Scher, A. Underwood, B. Jackson, V. Hill, J. T. McCrone, R. Colquhoun, C. Ruis, K. Abu-Dahab, B. Taylor, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution*, 7(2):veab064, 2021.

[46] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925, 2015.

[47] O. G. Pybus, A. Rambaut, and P. H. Harvey. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, 155(3):1429–1437, 2000.

[48] B. J. Quilty, L. A. Chapman, K. L. Wong, A. Gimma, S. Pickering, S. JD, R. P. G. Neil, C. I. Jarvis, and A. J. Kucharski. Reconstructing the secondary case distribution of SARS-CoV-2 from heterogeneity in viral load trajectories and social contacts. *Report for SPI-M-O and SAGE*, 2021.

[49] M. Ragonnet-Cronin, O. Boyd, L. Geidelberg, D. Jorgensen, F. F. Nascimento, I. Siveroni, R. A. Johnson, M. Baguelin, Z. M. Cucunubá, E. Jauneikaite, et al. Genetic evidence for the association between COVID-19 epidemic severity and timing of non-pharmaceutical interventions. *Nature Communications*, 12(1):1–7, 2021.

[50] A. Rambaut, E. C. Holmes, Á. O'Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, and O. G. Pybus. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature microbiology*, 5(11):1403–1407, 2020.

[51] D. Reich, R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. Johnson, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327):1053–1060, 2010.

[52] J. Riou and C. L. Althaus. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, 25(4):2000058, 2020.

[53] I. Smith and A. Ferrari. Equivalence between the posterior distribution of the likelihood ratio and a p-value in an invariant frame. *Bayesian Analysis*, 9(4):939–962, 2014.

[54] K. Sneppen, B. F. Nielsen, R. J. Taylor, and L. Simonsen. Overdispersion in COVID-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. *Proceedings of the National Academy of Sciences*, 118(14), 2021.

[55] A. Sobel Leonard, D. B. Weissman, B. Greenbaum, E. Ghedin, and K. Koelle. Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. *Journal of Virology*, 91(14):e00171–17, 2017.

[56] T. Stadler. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1):58–66, 2009.

[57] M. Steinrücken, A. Bhaskar, and Y. S. Song. A novel spectral method for inferring general diploid selection from time series genetic data. *The Annals of Applied Statistics*, 8(4):2203, 2014.

[58] K. Sun, W. Wang, L. Gao, Y. Wang, K. Luo, L. Ren, Z. Zhan, X. Chen, S. Zhao, Y. Huang, et al. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science*, 371(6526):eabe2424, 2021.

[59] R. N. Tasakis, G. Samaras, A. Jamison, M. Lee, A. Paulus, G. Whitehouse, L. Verkoczy, F. N. Papavasiliou, and M. Diaz. SARS-CoV-2 variant evolution in the United States: High accumulation of viral mutations over time likely through serial Founder Events and mutational bursts. *PloS ONE*, 16(7):e0255169, 2021.

[60] The COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. *The Lancet Microbe*, 1(3):e99, 2020.

[61] UK Health Security Agency. The R value and growth rate. https://www.gov.uk/guidance/the-r-value-and-growth-rate#:~:text=The%20R%20range%20for%20the,as%20of%2019%20March%202021.&text=The%20R%20range%20for%20the,as%20of%2012%20March%202021. Accessed: 2021-12-10.

[62] UK Office for National Statistics. Coronavirus (COVID-19) Infection Survey: England. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata. Accessed: 2021-12-10.

[63] E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, Á. O'Toole, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, 593(7858):266–269, 2021.

[64] E. M. Volz and X. Didelot. Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. *Systematic Biology*, 67(4):719–728, 2018.

[65] E. M. Volz, S. L. Kosakovsky Pond, M. J. Ward, A. J. Leigh Brown, and S. D. Frost. Phylodynamics of infectious disease epidemics. *Genetics*, 183(4):1421–1430, 2009.

[66] E. G. Williamson and M. Slatkin. Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, 152(2):755–761, 1999.

[67] T. Zinger, M. Gelbart, D. Miller, P. S. Pennings, and A. Stern. Inferring population genetics parameters of evolving viruses using time-series data. *Virus Evolution*, 5(1): vez011, 2019.

Chapter 4

Conclusion

The experimental and computational methods developed in this thesis allow us to infer the strength of genetic drift in a wider variety of settings. We developed a method to infer the strength of genetic drift in higher throughput in laboratory experiments of microbial range expansions. Using this method, we were able to test a fundamental question about genetic drift - how much variation in the strength of genetic drift results from random mutations? - as well as to study the phenotpyic and genetic determinants of genetic drift in range expansions. We found that single mutations can substantially change the strength of genetic drift and affect downstream evolution, suggesting that genetic drift can be an evolvable trait of a population. Additionally, we found that phenotypic traits that are emergent at the population level (such as colony shape, colony size, etc.) can explain most of the variation in genetic drift across genotypes.

We also developed a method to infer the strength of genetic drift while correcting for measurement noise using lineage frequency time-series data. Our method overcomes an important limitation of existing inference methods for genetic drift, which is the assumption of uniform sampling of data from the population. This method has allowed us to infer the strength of genetic drift in complex real-world data such as SARS-CoV-2 transmission dynamics in England. We find the levels of genetic drift in SARS-CoV-2 transmission in England are substantially higher than expected from the number of people infected with SARS-CoV-2 at the community level, even after correcting for measurement noise overdispersion and epidemiological dynamics. The levels of genetic drift that we detect are higher than estimates of superspreading from the literature. We find that a deme model of host contact structure can generate a high level of genetic drift even in the absence of superspreading. Our results suggest that further studies on heterogeneous host contact structures may be important for understanding the high levels of genetic drift observed for SARS-CoV-2 transmission in England.

4.1 Avenues for future work

Genetic drift as a potentially evolvable trait

There are many open questions and exciting avenues for future work. We have discovered that random mutations can substantially affect the strength of genetic drift in spatially-structured microbial colonies; while spatial-structured populations are very common in nature, it would also be interesting to study well-mixed populations, which also exist in nature and are more commonly used in microbial evolution experiments. For example, the *E. coli* Long Term Evolution Experiment (LTEE) is a well-mixed evolution experiment that has been running for over 70,000 generations, and has been extensively studied [4]. By studying how mutations affect genetic drift in well-mixed culture, we would be able to test whether genetic drift can evolve over time in the LTEE. If mutations do change genetic drift in well-mixed culture, what physiological changes lead to those changes? Some hypotheses are single cell lag time variation, death rate variation, and cell size variation.

Additionally, what are the consequences of genetic drift being an evolvable trait for evolutionary dynamics? Future theoretical work can explore how mutations that affect genetic drift ("drift modifiers") affect evolutionary dynamics in both well-mixed and spatially-structured populations. Relatedly, under what conditions would the drift modifiers that arise be sustained in the population? The evolutionary consequences of drift modifiers may be similar to (but not the same as) that of mutatators, which are mutations that change the mutation rate. Previous work has studied how mutators affect the rate of adaptation in well-mixed culture [1], which would be a helpful starting point for theoretical work on drift modifiers.

In our experiments, we focus on populations composed of a single genotype; however, most natural populations are composed of many genotypes that may have different phenotypic traits. What is the effect of drift modifiers in a population with many genotypes (with potentially different drift and fitness effects)? Relatedly, how do drift modifiers behave in a population with ecological dynamics in addition to evolutionary dynamics? These questions would be interesting to study using both theoretical and experimental approaches.

How do these processes change in microbial populations with more complex spatial structure [2]? For instance, biofilms have been shown to exhibit wrinkles, channels, delaminations, and streamers [5]. These higher-order structures may have interesting consequences for multiple evolutionary processes. New experimental methods may need to be developed to better understand evolution in biofilms, which are challenging to image with microscopy because they are often thick and opaque.

Genetic drift in pathogen transmission

In our analysis of genetic drift in pathogen transmission, we focused on SARS-CoV-2 in England because of the large number of available genomes. However, other countries such as the United States and Denmark have also collected a large number of genomes. It would

be interesting to apply the inference method developed in this thesis to study SARS-CoV-2 transmission in other countries and compare them to one another. Data from other countries that don't have as widespread of sequencing efforts may exhibit higher levels of measurement noise overdispersion than we observe in England; thus, the ability to correct for measurement noise overdispersion may be even more important for obtaining an unbiased measurement of the strength of genetic drift. It would also be interesting to apply the method developed here to study other pathogens with large amounts of genomic data, such as influenza.

In addition, our model of genetic drift in SARS-CoV-2 transmission only included genetic drift and measurement noise; in reality, other evolutionary and non-evolutionary processes are also occurring, such as natural selection, migration, mutation, recombination, and human behavioral changes. It will be interesting and important for future models to investigate the importance of these other processes to the observed population dynamics, and to incorporate them into inference models accordingly.

In our study, we focus on lineage frequency time series, but phylogenetic methods are highly developed and commonly used for studying pathogen dynamics. As both methods have their own benefits and drawbacks, future theoretical work should consider how to integrate these two types of methods when appropriate. For instance, how does the measurement noise parameter that we estimate in the lineage frequency time series method relate to the sampling parameter input that is used in phylogenetic birth-death models [7]? In this way, we can begin to leverage the benefits of both types of models to better understand pathogen dynamics, building on work that has already begun to tackle this challenge [3, 6].

4.2 Bibliography

[1] B. H. Good and M. M. Desai. Evolution of mutation rates in rapidly adapting asexual populations. *Genetics*, 204(3):1249–1266, 2016.

[2] J. Kayser, C. F. Schreck, Q. Yu, M. Gralka, and O. Hallatschek. Emergence of evolutionary driving forces in pattern-forming microbial populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1747):20170106, 2018.

[3] K. E. Kistler, J. Huddleston, and T. Bedford. Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. *Cell Host & Microbe*, 2022.

[4] R. E. Lenski. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME journal*, 11(10):2181–2194, 2017.

[5] A. Persat, C. D. Nadell, M. K. Kim, F. Ingremeau, A. Siryaporn, K. Drescher, N. S. Wingreen, B. L. Bassler, Z. Gitai, and H. A. Stone. The mechanical world of bacteria. *Cell*, 161(5):988–997, 2015.

[6] D. A. Rasmussen, O. Ratmann, and K. Koelle. Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Computational Biology*, 7(8):e1002136, 2011.

[7] T. Stadler. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1):58–66, 2009.