The SHAPE of logistic growth shows that timing does matter

Jonathan Dench[1,*],

1 Department of Biology, University of Ottawa, Ottawa, ON, Canada

* jdenc017@gmail.com

Abstract

Experimental evolution is a powerful tool for studying evolution and we see increasing use of *in silico* systems as our experimental questions go beyond what is practicable *in vivo/vitro*. However, existing *in silico* systems are limited by the models implemented, output obtainable, comparability of output from different software, or their lack of interpretability within biological systems. Here I introduce the R-package for SHAPE (Simulated Haploid Asexual Population Evolution - rSHAPE) a forward time simulation tool closely replicating microbial experimental evolution that offers the flexibility to implement alternative models and study detailed output. Thanks to a single simulation framework, rSHAPE users can better compare the results of diverse experimental conditions. While a modular programming approach makes the experimental conditions of rSHAPE readily expandable, at launch users may already tailor their experiments through more than 30 experimental parameters. While I suspect most users will be interested by the variable growth conditions and fitness landscape models implemented, certain users may appreciate that neither population nor genome sizes are not directed limited by rSHAPE but instead only by a users' computational resources. I validated the implementation of rSHAPE by replicating the expected evolutionary outcomes of several theoretical works strongly grounded in empirical observation. I then provide empirical evidence that the probability of a novel mutant surviving repeated rounds of serial passaging is conditional upon when during a growth phase the mutant arises in a

community that undergoes logistic growth. Unlike the roughly uniform survival probability predicted by existing theory, which only considered exponential growth, I find that the timing of mutant arrival matters when growth is logistic. Users may download software and documentation for rSHAPE through CRAN mirrors or via GitHub at `https://www.github.com/Jdench/SHAPE_library/`.

Author summary

The forward time simulation tool rSHAPE (R-package for Simulated Haploid Asexual Population Evolution) was designed to compliment the study of both theoretical and empirical evolution. Included with rSHAPE are functions to programmatically build, run, and initially process results of an evolutionary experiment defined by a range of parameters. As experimental evolution often studies both the emergence and fate of *de novo* mutants, I validated rSHAPE by confirming its ability to replicate seminal theoretical expectations concerning changes in fitness through time and the fixation probability of mutants. I then applied rSHAPE to study how the laboratory protocol of serial passaging, common in microbial experimental evolution, affects the fixation probability of *de novo* mutants. Unlike related theoretical work which modelled growth as effectively exponential [1], this study considered populations undergoing logistic growth which is common to microbes undergoing serial passaging. In contrast to exponential growth, when a community undergoes logistic growth the probability of a mutant arising and eventually fixing depends upon when during a growth phase a mutant arises.

Introduction

While certain drivers of evolutionary dynamics have been elucidated, researchers are increasingly interested in studying the fate of populations, and the mutations they carry, in a more wholistic context [2]. While experimental evolution can be paired with next generation sequencing technologies in order to track the emergence and fate of mutations [3], these *in vivo/vitro* approaches are limited by physical or financial costs. Even when experimental design is optimised, it is still not feasible for most researchers

to fully sequence whole genomes, of whole populations, at multiple intervals, for each replicate of an experiment. This proposition is not fanciful idealism as the interaction among mutations can affect evolutionary outcome [4] and at least one research group has effectively realised this dream [5]. While evolutionary theory and analytical analyses are much less costly approaches, we need to combine both evolutionary theory and empirical study to improve our understanding of evolution in increasingly complex systems [6]. Simulations, or *in silico* evolution, can leverage evolutionary theory based on empirical evidence to support both theoretical and empirical study of evolutionary biology, while being limited only by the available software and computational resources.

Simulation tools complement theoretical and empirical study by allowing various theoretical models, and combinations of the underlying evolutionary parameters, to be tested through comparison of simulation results and empirical observations. There exist simulation tools to track changes in population demographics [7] or specific mutations [8–11], though most of these implement a single theoretical framework and provide only specific output. The AVIDA [12] and AEVOL [13] tools are remarkable simulation platforms for studying genetic changes but their evolutionary systems are abstractions that are practical for their implementation but make it unclear to what extent their results apply to living systems. A powerful tool for simulating microbial experimental evolution is the Haploid Evolutionary Constructor (HEC) [14], though it simplifies evolution of the genome by mapping each locus to a single quantitative trait value and assumes mutations do not interact. Knowing no single framework to simulate experimental evolution while tracking detailed changes in both population demographics and all mutations in large genomes (*i.e.*: millions of mutational sites), I found that comparing different evolutionary scenarios simulated by the variety of existing software would likely be impractical even when possible. To address this, I developed the R-package for SHAPE (Simulated Haploid Asexual Population Evolution - rSHAPE).

The simulation tool rSHAPE was designed to simulate microbial experimental evolution in a framework that can readily be expanded to different theoretical models thanks to a modular programming design. I chose to simulate microbial evolution experiments as they are commonly used for the manipulative study of evolution. Microbial evolution experiments are often used to study how factors influence the emergence and persistence of *de novo* mutants arising from an initially isogenic

population. The design of many microbial evolution experiments requires the practice of serial passaging, a protocol in which a small fraction of stationary phase communities are transferred to fresh growth media to extend the evolutionary time of experiments. When communities are well mixed prior to serial transfer, population loss is proportional to their representation in the community. While serial transfer is expected to result in the loss of rare mutants, theory suggests that the probability of a mutant arising during exponential growth, and subsequently surviving many rounds of serial transfer, is roughly uniform [1]. It is only in chemostats that microbial populations continuously grow exponentially, but chemostat setups are often avoided because they allow selection for biofilm development [15] which is often not desirable. So, most experimental microbial communities will undergo logistic growth between serial transfers and it remains unclear how serial transfer may affect the joint probability of mutants arising and subsequently fixing.

In this paper I begin by explaining the framework of rSHAPE. I then report on rSHAPE's ability to reproduce expected trends in fitness through time as well as the probability of a *de novo* mutant fixing under the conditions of several theoretical models. I then apply rSHAPE to test if serial passaging influences the fixation probability of a *de novo* mutant arising in a community that undergoes logistic growth.

Design and implementation

Simulating evolution

rSHAPE is a forward time, discrete step, *in silico* experimental evoltion software that tracks demographic changes for the distinct populations of haploid genotypes that form an evolving community. Haplotype genomes are tracked as their ordered set of L binary state sites (*i.e.* genome of length L) where 0 represents the unmutated state and 1 depicts mutation. Note that while it may be practical to consider each site as a nucleotide base, each site could equally be considered as any mutable element which influences genotype fitness (*e.g.* a codon or gene). To define the experimental conditions, rSHAPE includes the function *defineSHAPE* which allows users to directly set 30 experimental parameters (Appendix A, Table S1). Once the parameters are

defined, the experiment is started by calling the function *runSHAPE* which will sequentially simulate the evolution of n independently replicated starting communities. In each time step, individuals may die, reproduce, generate mutants, and experience stochastic loss events (*e.g.* a population bottleneck caused by serial transfer). While it may be practical to think of a time step as a biological generation in which each individual of average fitness will produce r offspring (where r is the growth rate, often modelled as 2 for bacteria), this need not be the case (*e.g.* users can set the average probability of birth such that $P_{birth} \in [0, 1]$). Regardless, after each time step rSHAPE records the evolving communities' demographics as the current size of each population and their numbers of deaths, births, mutants generated (decreases population size), as well as newly arisen mutants (increases population size) including the "parent" population from which they arose. By tracking the progenitor of mutants, the entire evolutionary history can be reconstructed permitting detailed analyses of evolutionary trajectories. To further support evolutionary analyses, rSHAPE tracks the full fitness landscape of all explored mutational space. Explored mutational space is considered to be the possible mutant genotypes produced by any "parent" genotype. At the onset of a run of rSHAPE, the evolving community will consist of a single population with K wild-type (0 state for all genomic positions) individuals. In its current form, rSHAPE resembles many microbial evolution experiments in that populations are finite and evolve in a single, well mixed (*i.e.* homogeneous) environment. While these are conditions at the launch of rSHAPE, the starting community composition, evolutionary environment, as well as other implemented evolutionary models (*e.g.* growth forms, fitness landscapes) could readily be changed due rSHAPE's modular programming design. This modular design is realized by each run of rSAHPE simply being a set of establishing function calls followed by discrete time steps which are iterative function calls that control the behaviour of stochastic events, deaths, births, and mutation (Fig. 1). So while in this paper I will describe the current state of rSHAPE at launch, its modular framework simplifies feature changes. For a detailed description of the iterative functions called during a run of rSHAPE, please refer to Appendix S1.

Running experiments

While populations simulated with rSHAPE are finite, I have set no upper bound on either their size or their genome length. It is possible to define an evolving community of individuals with genomes containing trillions of positions and which grows exponentially without death. Though this would be a possible extreme use case, from early testing of rSHAPE I identified serious random access memory (RAM) issues when simulating realistic microbial populations containing billions of individuals with genomes having millions of positions. To mitigate this challenge, rSHAPE exports the full stepwise demographic breakdown, as well as the explored fitness landscape of all genotypes, to SQL databases which act as both the experimental record and a reference when needed during the run. As a result, *in silico* evolution with rSHAPE is less limited by RAM than by disk space and the number of available processors to run independent replicates and parameterizations.

To facilitate rSHAPE's use for simulating a breadth of experimental conditions, I have included the *shapeExperiment* function that uses template files found here: `https://github.com/JDench/SHAPE_library/tree/master/SHAPE_templates`. With *shapeExperiment*, rSHAPE will build parameter combinations from user defined values saved in the template control files and create an experimental folder within which all files necessary to run the experiment will be written (these include bash shell batch submission scripts). Before ending each independently replicated simulation, rSHAPE will generate an RData summary of the run to facilitate analysis of the results. Further, once all simulations in an experiment are complete, users can use the *summariseExperiment* function which will generate a summary of the RData summaries to further facilitate analyses. Experiments built with *shapeExperiment* will have a script to run *summariseExperiment*. At present, *summariseExperiment* will create output to facilitate comparison of population dynamics, evolutionary trajectories, and the repeatability of evolution but future improvements will expand the breadth of summary analytics and allow users to more easily select from among the set of possible summary output.

Results and discussion

Validating implementation

To validate the implementation of rSHAPE, I assessed its ability to simulate expected evolutionary trends as well as replicate the evolutionary dynamics predicted by seminal theoretical works. Since the goal of many microbial evolution experiments is to analyse factors that affect the emergence and maintenance of mutations, I first tested that communities evolving in rSHAPE show expected fitness trends, and second that evolutionary dynamics simulated with rSHAPE can replicate the theoretical fixation probability of *de novo* mutants under different growth conditions.

To test rSHAPE's ability to replicate standard evolutionary trends, I initially tested that the fitness of evolving communities increase through time and that population bottlenecks alter the rate of fitness increase [1] (Fig. 2). For this, I simulated large communities (carrying capacity $K = 1,000,000$) of individuals with 100 genomic positions ($L = 100$), evolving for $T = 10,000$ generations, for each of three different regular bottleneck sizes ($D \in 10, 100, 1000$), where mutant genotype fitness was calculated with either an additive model (a smooth fitness landscape defined by independent mutation fitness contributions - *i.e.* no interaction between mutations, or no epistasis) or a House of Cards (HoC) fitness landscape [16, 17] (the fitness of every genotype is independently drawn from the same distribution - *i.e.* a maximally rugged landscape). The independent effect of all mutations was assumed to be beneficial and so based on extreme value theory [18], the random component of all genotype fitness calculations was drawn from an exponential distribution ($\lambda = 100$). For statistical comparison, I replicated each parameter combination 500 times by initialising five independent fitness landscapes within which evolution was replicated 100 times. For the additive fitness landscape models, I expected fitness to increase until all mutations had fixed since all mutational effects were beneficial. However, each genotype simulated with a HoC fitness landscape is equally likely to be the global optimum and many local optima are expected to exist [17], so that fitness increase should generally stop long before all mutations fix in a single population. From analysis of the evolutionary trends in these first simulations, I found for the additive fitness landscape models that both fitness and the number of mutations increased throughout because the "full" mutant

was never reached, whereas in the HoC models evolution stopped after only a few mutations had fixed and a local optima had been found (SI Fig. 3).

As my second assessment, I tested if rSHAPE could replicate seminal theoretical works related to the fixation probability of *de novo* mutants. The theoretical models considered here only have closed form approximate solutions however, using a search of parameter space allows us to find an exact theoretical probability of fixation for the simplest case of communities with constant size [19]. Further, all the theoretical works considered here are built upon the initial modelling of the simplest case when community size is constant [20–22]. For these reasons, I expected rSHAPE to effectively replicate theoretical fixation probabilities when assumptions of the models' approximate solutions were not violated. From tests, I found that rSHAPE was able to replicate the predicted fixation probability of *de novo* mutants arising in communities of constant size [19] (Fig. 4, and Appendix S2: "Haldane's $2s$") as well as for those experiencing logistic growth [20, 21] (Fig. 5, and Appendix S2: "Logistic Growth"). In these comparisons, rSHAPE was able to exactly replicate the theoretically predicted results provided the assumptions of their analytical approximations were not violated (Figs. 4A, 5A). Under parameterisations which violated the underlying assumptions (*i.e.* high growth rate and/or mutational selection coefficient), rSHAPE underestimates fixation probability by an amount proportional to the difference between the exact probability of fixation and the analytical approximation for communities of constant size [19] (*i.e.* exact theoretical probability *vs* Haldane's $2s$ approximation; Figs. 4B and 5B). Thus, I have found that the evolutionary conditions implemented in rSHAPE can replicate theoretically expected evolutionary dynamics under proper parameterizations. Further, from testing parameterizations that violated model assumptions I found that rSHAPE deviates from approximate solutions by an amount proportional to the difference between the exact and approximate solutions of the underlying model. I suggest that rSHAPE provides a means to test evolutionary scenarios under conditions not supported by existing analytical models.

All simulation scripts, and resulting summary files, can be found at `github.com/Jdench/rSHAPE_validation`. The full results for this section are available upon request.

rSHAPE application

Having validated the implementation of rSHAPE, I applied the simulation framework to assess the fixation probability of *de novo* mutants arising in a community that undergoes logistic growth and periodic population bottlenecks. It is common for microbial experimental evolution to include serial passaging, a practice that allows continued evolution by transferring a portion of stationary phase communities to fresh growth media. The process of serial passaging introduces periodic bottlenecks which delineate growth phases of the community and is expected to bias the types of successful mutations [23]. Earlier theoretical work suggests that the probability of a mutation both arising within the community and surviving repeated rounds of serial passaging is roughly uniform throughout growth phases that are effectively exponential [1]. However, there are no estimates of how serial passaging affects fixation probability when growth is logistic in these communities.

To estimate this, I used rSHAPE to simulate evolution of communities undergoing repeated rounds of serial passaging after growth phases that were modelled as either exponential or logistic growth. To compare the results of rSHAPE with previous theoretical work, I parameterized my simulations following the exponential and nutrient limited growth conditions of *Wahl et al.* (2002). The nutrient limited growth parameters were applied to my simulations of logistic growth. I found that rSHAPE does estimate a nearly uniform survival probability for *de novo* mutants arising in an exponential growth community (Fig. 6A). The parameters used in these simulations included a growth rate ($r = 2$) and mutation selection coefficient ($s = 0.1$) that would be deemed as "high" values outside parameter range of the simplifying assumptions in antecedent theoretical works [19, 21]. Similar to my findings when validating the implementation, I had expected rSHAPE to estimate a lower fixation probability than the analytical approximation of *Wahl et al.* (2002) which it did (Fig. 6A). For conditions of logistic growth, I found that the joint probability of a mutant arising during the growth phase and subsequently surviving repeated rounds of serial passaging was not uniform (Fig. 6B). To better understand what evolutionary factors contributed to this result, I visualized and compared the individual probabilities that the mutant lineage arose during the growth phase, ultimately survived serial passaging, and also measured the

expected mutant lineage size by the end of the growth phase in which it arose (Fig. 6C,D,E respectively). Note that the joint probabilities estimated for exponential and logistic growth differ by two orders of magnitude (Fig. 6A,B), this reflects the magnitude difference between the product of the initial populations size and mutation rate parameters used ($N_0 \in (10^5, 10^7)$ and $\mu \in (5 \times 10^-5, 4 \times 10^-9)$ for the exponential and logistic growth conditions respectively). This order of magnitude difference is reflected in the estimated probability of the mutant's arrival throughout a growth phase (Fig. 6C). Under logistic growth, *de novo* mutants have a relatively higher probability of arising at all times up until the environmental carrying capacity is nearly reached (Fig. 6C). Yet, the probability of surviving repeated rounds of serial passaging is lower when growth is logistic (Fig. 6D) an observation likely underlain by the relatively smaller mutant lineage population size achieved in the first logistic growth phase (Fig. 6E).

While a simple application, this example demonstrates how rSHAPE can be used to test hypotheses under more complex evolutionary scenarios where analytical modelling is not yet practicable. By applying rSHAPE to study the effects of serial passaging on *de novo* mutants arising during logistic growth, I provide evidence that the timing of mutant arrival during a growth phase does matter. If we assume that any, but not all, "next evolutionary step" mutants are equally likely to result from mutations events throughout the growth phase, the biased fixation probability relative to time of arrival will increase stochasticity between replicates. Since many microbial evolution experiments include serial passaging within their experimental design, my findings suggest that serial passaging may reduce the repeatability of genomic changes observed in those experiments.

Availability and future directions

rSHAPE is made available through CRAN or can be installed from GitHub `github.com/JDench/SHAPE_library`. This software is made freely available under the GNU General Public License v3.0.

Future developmental priorities for rSHAPE are threefold. First, I would expand the applicability of rSHAPE by including methods to handle different habitat patches, partitioning of the genome, and temporal heterogeneity to selection. Second, implement

additional fitness landscape models such Fisher's Geometric Model [24] and an eggbox model [25]. Lastly, I want to translate rSHAPE from the language of R to C in order to improve runtime through use of POSIX thread libraries and ultimately reducing time for SQL database queries.

Acknowledgments

I would like to thank the Center for Advanced Computing (CAC) for access to their computational resources. I would also like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada for funding support accessed through the grants of my supervisors Drs. Stéphane Aris-Brosou and Rees Kassen.

Supplementary Information

S1 Appendix: Detailed information concerning the iterative functions that correspond to a run of rSHAPE.

S2 Appendix: Detailed information concerning comparison of rSHAPE to seminal theoretical works studying the fixation probability of *de novo* mutants.

References

1. Wahl LM, Gerrish PJ, Saika-Voivod I. Evaluating the impact of population bottlenecks in experimental evolution. Genetics. 2002;162(2):961–971.

2. Losos JB, Arnold SJ, Bejerano G, Brodie ED III, Hibbett D, Hoekstra HE, et al. Evolutionary Biology for the 21st Century. PLOS Biology. 2013;11(1):1–8. doi:10.1371/journal.pbio.1001466.

3. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics. 2016;17(6):333–351. doi:10.1038/nrg.2016.49.

4. Phillips PC. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9(11):855–867.

5. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. The dynamics of molecular evolution over 60,000 generations. Nature. 2017;doi:10.1038/nature24287.

6. de Visser JAGM, Krug J. Empirical fitness landscapes and the predictability of evolution. Nat Rev Genet. 2014;15(7):480–90. doi:10.1038/nrg3744.

7. Guillaume F, Rougemont J. Nemo: an evolutionary and population genetics programming framework. Bioinformatics. 2006;22(20):2556–2557. doi:10.1093/bioinformatics/btl415.

8. Lambert BW, Terwilliger JD, Weiss KM. ForSim: a tool for exploring the genetic architecture of complex traits with controlled truth. Bioinformatics. 2008;24(16):1821–1822. doi:10.1093/bioinformatics/btn317.

9. Dalquen DA, Anisimova M, Gonnet GH, Dessimoz C. ALF—A Simulation Framework for Genome Evolution. Molecular Biology and Evolution. 2012;29(4):1115–1123. doi:10.1093/molbev/msr268.

10. Zanini F, Neher RA. FFPopSim: an efficient forward simulation package for the evolution of large populations. Bioinformatics. 2012;28(24):3332–3333. doi:10.1093/bioinformatics/bts633.

11. Arnold BJ, Gutmann MU, Grad YH, Sheppard SK, Corander J, Lipsitch M, et al. Weak Epistasis May Drive Adaptation in Recombining Bacteria. Genetics. 2018;doi:10.1534/genetics.117.300662.

12. Beckmann BE, Clune J, Ofria C. Digital Evolution with Avida. In: Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation. GECCO '10. New York, NY, USA: ACM; 2010. p. 2917–2926. Available from: `http://doi.acm.org/10.1145/1830761.1830923`.

13. Knibbe C, Coulon A, Mazet O, Fayard JM, Beslon G. A long-term evolutionary pressure on the amount of noncoding DNA. Mol Biol Evol. 2007;24(10):2344–2353.

14. Lashin S, Matushkin YG, Klimenko A, Suslov VV, Kolchanov NA. Evolution, Biodiversity and Ecology in Microbial Communities: Mathematical Modeling and

Simulation with the "Haploid Evolutionary Constructor"Software Tool. IntechOpen; 2014.

15. Larsen DH, Dimmick RL. ATTACHMENT AND GROWTH OF BACTERIA ON SURFACES OF CONTINUOUS-CULTURE VESSELS. Journal of Bacteriology. 1964;88(5):1380–1387.

16. Kingman JFC. A Simple Model for the Balance between Selection and Mutation. Journal of Applied Probability. 1978;15(1):1–12.

17. Kauffman S, Levin S. Towards a general theory of adaptive walks on rugged landscapes. Journal of Theoretical Biology. 1987;128(1):11 – 45. doi:http://dx.doi.org/10.1016/S0022-5193(87)80029-2.

18. Gillespie JH. Molecular Evolution Over the Mutational Landscape. Evolution. 1984;38(5):1116–1129.

19. Haldane JBS. A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. Proceedings of the Cambridge Philosophical Society. 1927;23:838. doi:10.1017/S0305004100015644.

20. Ewens W. The probability of survival of a new mutant in a fluctuating environment. Heredity. 1967;22(3):438–443.

21. Otto SP, Whitlock MC. The Probability of Fixation in Populations of Changing Size. Genetics. 1997;146(2):723–733.

22. Wahl LM, Gerrish PJ. The Probability That Beneficial Mutations Are Lost in Populations with Periodic Bottlenecks. Evolution. 2001;55(12):2606–2610.

23. Wahl LM, Zhu AD. Survival Probability of Beneficial Mutations in Bacterial Batch Culture. Genetics. 2015;200(1):309–320. doi:10.1534/genetics.114.172890.

24. Fisher RA. The Genetical Theory of Natural Selection. 2nd ed. New York: Dover Publications; 1958.

25. Ferretti L, Schmiegelt B, Weinreich D, Yamauchi A, Kobayashi Y, Tajima F, et al. Measuring epistasis in fitness landscapes: The correlation of fitness effects

of mutations. Journal of Theoretical Biology. 2016;396:132 – 143. doi:http://dx.doi.org/10.1016/j.jtbi.2016.01.037.

Figures

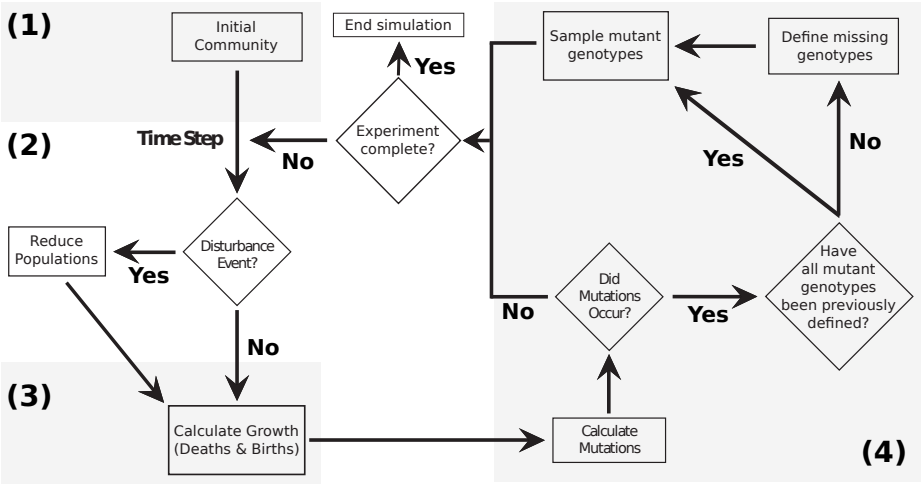

Fig 1. rSHAPE's flowchart. The calculation of disturbance events, deaths, births and mutations are performed for each of the populations with at least one living member prior to the function call. Deaths and births are treated as occurring simultaneously and handled by the wrapper function for growth. Experiments end either when the defined number of evolutionary steps have been simulated or there are no living populations in the community.

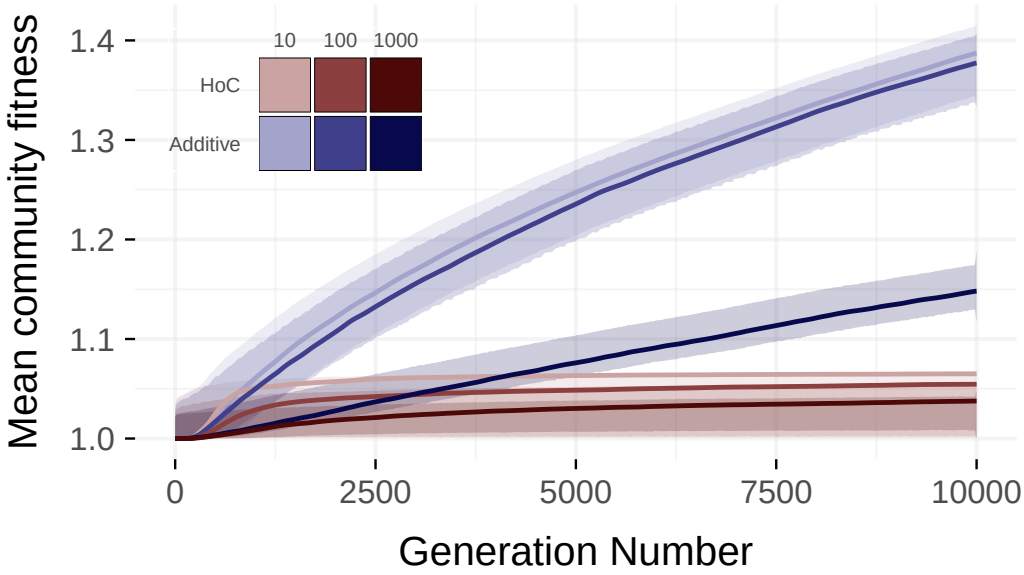

Fig 2. Community fitness through time simulated with rSHAPE. Colors represent results from different fitness landscape models (red - additive, blue - House of Cards, HoC) and shading reflects intensity of regular population disturbance events. Solid lines represent the trend in mean community fitness across all replicates while the surrounding shaded polygon represent the absolute minimum and maximum values realised during simulations.

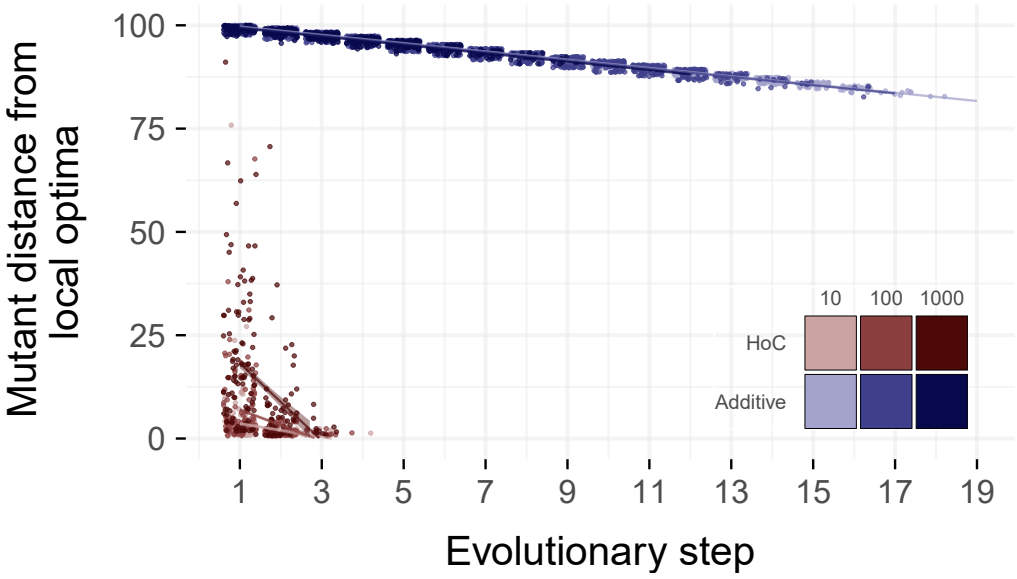

Fig 3. Distance of successful mutants from local optima as a function of ordered evolutionary step. Colors represent results from different fitness landscape models (red - additive, blue - House of Cards, HoC) and shading reflects intensity of regular population disturbance events. Solid lines represent the average trend of mutational distance between "next evolutionary step" mutants and a local fitness optimum. Mutational distance represents the number of single mutations that separate two genotypes.

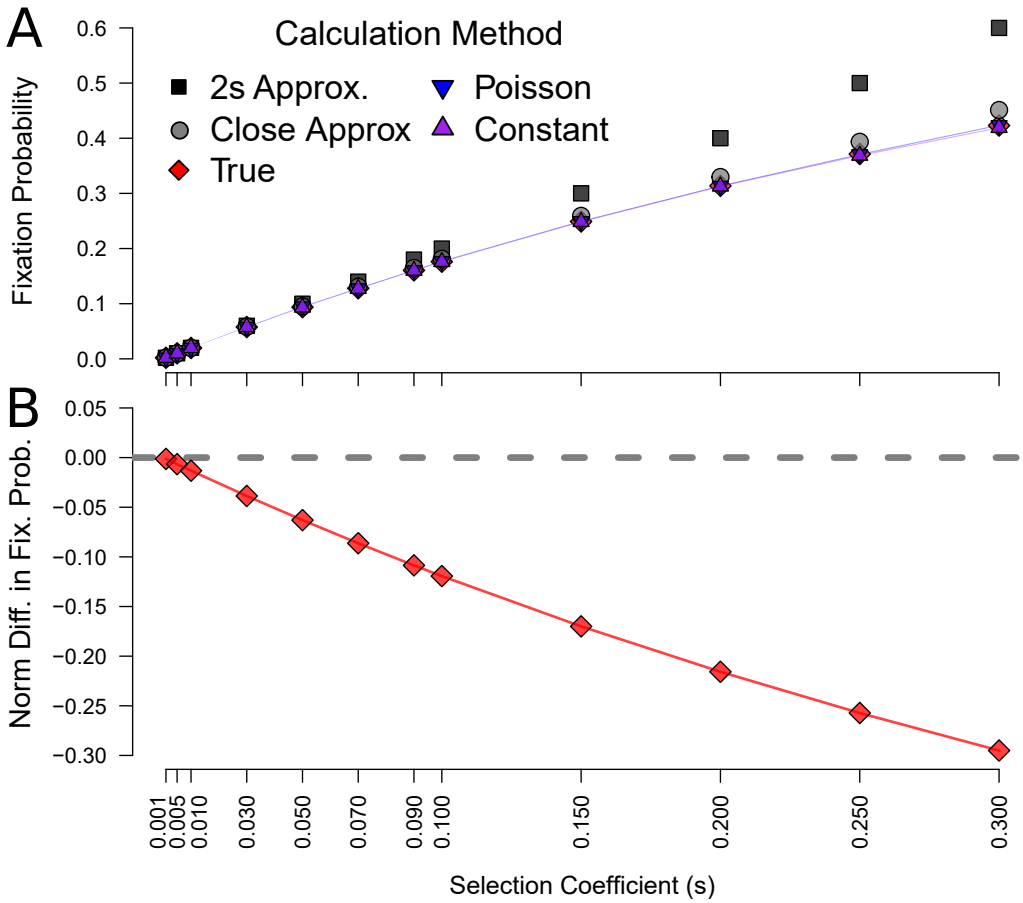

Fig 4. Comparison between the probability of a single mutant fixing, in an environment with a constant number of individuals, as calculated with rSHAPE or using a theoretical model [19]. A shows the fixation probability, dependent on the selection coefficient s, where coloured circles represent the values calculated using different constant growth models (*Poisson* - blue; *Constant* - purple) implemented in rSHAPE, red diamonds represent the true fixation probability, black squares *Haldane's 2s* approximation and grey circles the closer approximation $1 - e^{-2s}$. Note, the red diamonds are not always evident as they are overlain by the fixation probabilities estimated with rSHAPE. B, I present the normalised difference between the true fixation probability and the analytical approximation of $2s$. Negative values reflect that the true fixation probability is lower than the analytical approximation of $2s$.

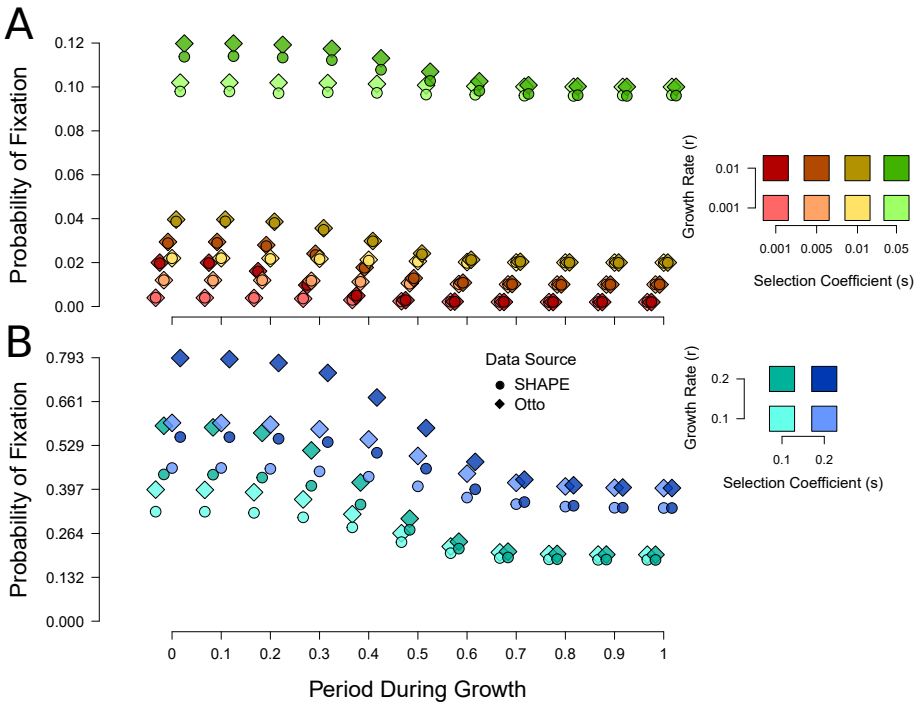

Fig 5. Comparison of the theoretical and estimated fixation probabilities for a mutant growing logistically. Diamond shapes represent the analytical approximation [21] derived similarly to Haldane's $\approx 2s$ while circles are for estimates using rSHAPE. The color reflects the selection coefficient (s) with darker colours represent higher intrinsic growth rates (r). The period during the growth phase is scaled from the start of growth until the point where the number of individuals reaches carrying capacity. A presents the range of parameters originally studied in the theoretical work whereas B shows larger parameter values.

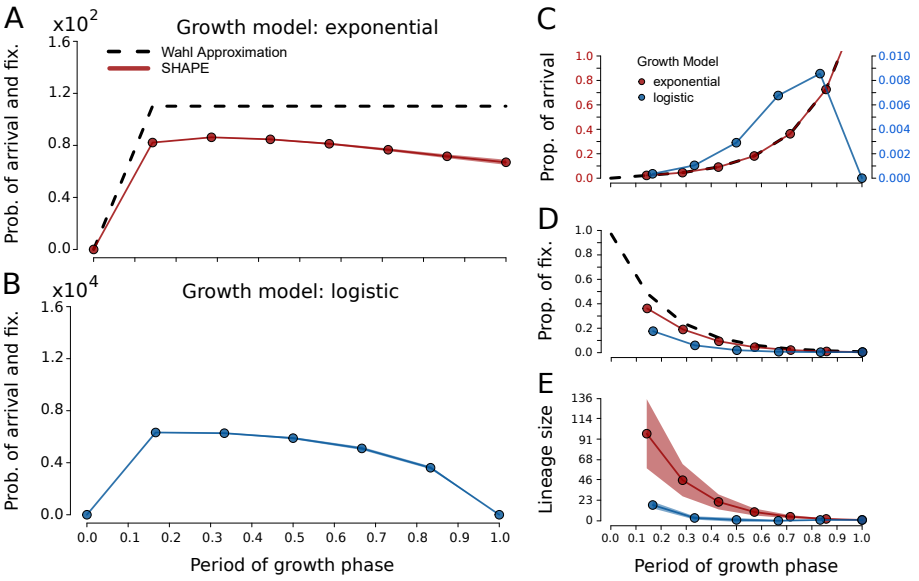

Fig 6. The joint probability of a single mutant arising during growth and then surviving many rounds of community bottlenecks. The parameters used are similar to those of Wahl *et al.* [1] such that the mutation has a selection coefficient $s = 0.1$, bottleneck strength was set to $D = 100$, and other parameters varied based on the model of growth (see main text). A compares the discrete analytical approximations of Wahl *et al.* (2002), and the estimates of rSHAPE, for the joint probability of mutant fixation in a community growing exponentially. B shows the probability calculated from simulations with rSHAPE when growth was logistic. C shows the independent probabilities of a mutant having arisen at a time during the growth phase whereas D visualises the probability to have ultimately survived repeated rounds of serial passaging. E shows estimates of the mutant lineage size by the end of the first growth phase in which it arose.