STUN: forward-time **S**imulation on **TU**nable fit**N**ess landscapes in recombining populations

André Amado[1,2,*,a], Juan Li[1,2,*,a], and Claudia Bank[1,2]

[1]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland
[2]Swiss Institute for Bioinformatics, Lausanne, Switzerland
[*]These authors contributed equally to this work.
[a]To whom correspondence should be addressed: andreamado1@gmail.com, juan.li@unibe.ch

Abstract

Simulation is a powerful tool to study polygenic trait adaptation under the influence of complex evolutionary forces. Here, we implement a forward-in-time population-genetic simulation package based on Wright-Fisher dynamics. In this package, we assume that haploid recombining populations use standing genetic variation to adapt to a sudden environmental change, exposing the population to a new fitness landscape. The fitness landscapes can contain any level of epistasis. We offer out-of-the-box implementations for five fitness landscape models. In addition, our package allows for the specification of any custom fitness landscapes and recombination maps.

Availability and implementation: STUN is implemented in Rust. Its source code is available at `https://github.com/banklab/STUN`, which also includes a link to the software's manual and binary files for Linux, macOS and Windows.

1 Introduction

How does a population adapt upon sudden environmental change? Adaptation often happens through changes in polygenic traits. Many studies have approached the genetic basis of such traits empirically and theoretically. However, identifying the genetic basis and its adaptive

consequences is challenging because of the vastness of the underlying genotype space. In this space, evolution is difficult to predict, and therefore the adaptive path can be impossible to disentangle or recover. In this context, computational approaches play a crucial role in exploring potential adaptation scenarios. Few simulators exist that are efficient and flexible enough to incorporate recombining populations that adapt to complex fitness landscapes. When we consider non-additive effects (epistasis and dominance) on a fitness landscape, their interplay requires complicated assumptions and causes a large computational burden. Hence, the common simulators (e.g., simuPOP and SLiM) focus on simulating direct, non-epistatic selection in diploid populations (Peng and Kimmel [2005], Haller and Messer [2019]). These packages provide straightforward means to define dominance at a locus, but they do not allow for the straightforward implementation of ubiquitous and complex epistasis as is expected in nature (Bank [2022]).

We built a software named STUN to address the relative contribution of direct selection and epistasis to adaptation. Our software simulates the adaptation of haploid populations with recombination evolving on arbitrary fitness landscapes (Figure S1). In this software, the fitness landscape generator provides an out-of-the-box implementation of five common fitness landscape models. The five models are the Rough Mount Fuji, the NK, the Block, the additive and the House-of-Cards model (reviewed in Fragata et al. [2019]). The last two models represent the limiting cases of the previous ones. Besides the built-in models, our software supports arbitrary user-defined fitness landscapes.

The initial step of the simulation is to generate a population by specifying initial allele frequencies, either by allocating the same allele frequencies to every locus or by drawing from a folded neutral site frequency spectrum (SFS, Hudson [2015]). For the neutral SFS, it is possible to set a genetic drift threshold to exclude low-frequency alleles not experiencing significant selection during the early generations (see Li et al. [2023]). The reproduction step is simulated by a Wright-Fisher model with non-overlapping, discrete generations. Within each generation, the individuals are randomly paired and perform recombination to generate a gamete pool. The recombination map can be uniform by assuming that the recombination probability is constant between neighboring loci. Alternatively, any custom recombination map can be specified. Subsequently, haploid offspring are sampled from the gamete pool with a probability proportional to their fitness to form the new generation. This process continues until all loci are monomorphic.

Detailed data about the population and adaptation process are collected at the user's discretion. A customizable suite of statistics is available, *e.g.*, mean population fitness or haplotype diversity. This population information can be recorded every nth generation. If $n = 1$, the whole adaptive trajectory is recorded. At the end of each adaptive process, summary statistics are saved, *e.g.*, fixation time of each allele or the population fitness distance to the global peak. The user can also output the full population at any generation. Details about data

and available statistics are further explained in the user manual (see Supplementary Material).

We have previously applied this software to investigate how the interplay between recombination and epistasis affects adaptation (Li et al. [2023]). To illustrate the usefulness of our program, we give an additional toy example here. We simulated populations that adapt to Mount Rough Fuji fitness landscapes with three levels of epistasis. We extracted the variance in population fitness at each generation to illustrate that a high level of epistasis confers rapid fixation (Figure S2). Consistent with Fisher's fundamental theorem of natural selection, the rate of adaptation is magnified by the large variance in population fitness caused by epistasis, leading to a shorter fixation time.

2 Performance

The sizes of the fitness landscape and of the population have a large effect on software performance. The mean computation time increases with the sizes of the fitness landscape and of the population. To quantify this, we measured the mean computation time per generation as a proxy for performance. We show the computation time for three different recombination rates, and compare results for all or no statistics. Figure 1a shows the dependence of performance on the number of loci. The computation time per generation scales linearly with the number of loci (Table S1) despite the exponential growth of the number of genotypes. The population size affects the computation time in a more complex way (Figure 1b). A linear fit does not capture the entire range of data, and an additional quadratic term is required (Table S1). This quadratic term is always negative, indicating a less-than-linear increase in the computation time with the population size. We can see from both panels that the calculation of statistics does not add a substantial burden to the calculation time, except when small population sizes are combined with no recombination. Further details about the performance measurements are provided in the user manual.

3 Conclusion

STUN allows for efficient simulation of rapid adaptation on arbitrary fitness landscapes. By including recombination and standing genetic variation, STUN captures the complex dynamics of adaptation of recombining organisms. In addition, we can use existing empirical and theoretical fitness landscapes in simulations with STUN to develop expectations of complex evolutionary dynamics, which can guide, for example, the experimental design of evolve-and-resequence experiments.

Acknowledgements

We thank the THEE lab members for suggestions that improved this work.

Funding

This work was supported by funding from ERC Starting Grant 804569 (FIT2GO), SNSF Project Grant 315230/204838/1 (MiCo4Sys), and HFSP Young Investigator Grant RGY0081/2020 to CB.

References

C. Bank. Epistasis and adaptation on fitness landscapes. *Annual Review of Ecology, Evolution, and Systematics*, 53, 2022.

I. Fragata, A. Blanckaert, M. A. D. Louro, D. A. Liberles, and C. Bank. Evolution in the light of fitness landscape theory. *Trends in ecology & evolution*, 34(1):69–82, 2019.

B. C. Haller and P. W. Messer. Slim 3: forward genetic simulations beyond the wright–fisher model. *Molecular biology and evolution*, 36(3):632–637, 2019.

R. R. Hudson. A new proof of the expected frequency spectrum under the standard neutral model. *Plos one*, 10(7):e0118087, 2015.

J. Li, A. Amado, and C. Bank. Rapid adaptation of recombining populations on tunable fitness landscapes. *bioRxiv*, 2023. doi: $10.1101/2022.11.16.516770$.

B. Peng and M. Kimmel. simupop: a forward-time population genetics simulation environment. *Bioinformatics*, 21(18):3686–3687, 2005.

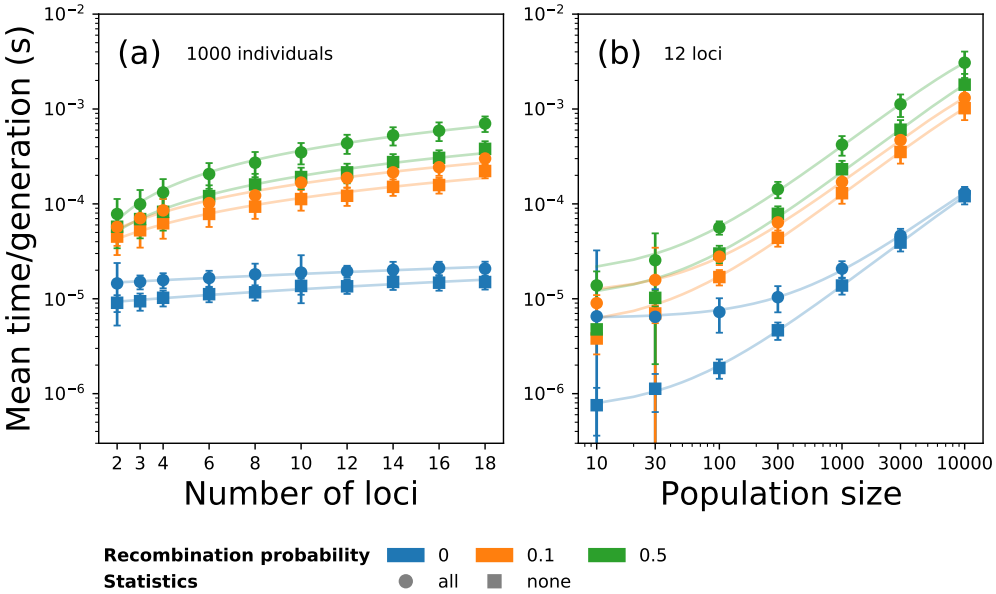

Figure 1: Performance. We measured the mean computation time (in seconds) per generation (dots). The bars represent one standard deviation. The underlying curves show polynomial fits to each data set. A. Mean time per generation as a function of the number of loci in the fitness landscape. The fit is linear. B. Mean time per generation as a function of the population size. The fit is a second-order polynomial. All simulations were performed on Rough Mount Fuji fitness landscapes, with additive and epistatic effects drawn from normal distributions $\mathcal{N}(0, 0.01)$. We simulated adaptive walks over 100 landscapes, and 10 populations for each fitness landscape. All initial populations were sampled from a neutral site frequency spectrum. The performance was measured in two conditions, either including none (circles) or the complete set of statistics available in the package (squares). All simulations were run on a 2.3 GHz Intel Core i9 CPU.