Predicting the Unpredictable: Using replay experiments to disentangle how evolutionary outcomes are altered by adaptive momentum

Austin J. Ferguson[1,2,3,*], Charles Ofria[1,2,3], Clifford Bohm[3,4]

[1]Department of Computer Science and Engineering [2]Program in Ecology, Evolution, and Biology
[3]BEACON Center for the Study of Evolution in Action [4]Department of Integrative Biology
Michigan State University, East Lansing, MI 48824
[*]fergu358@msu.edu

Abstract

When a new evolutionary dynamic is identified, researchers often struggle to understand its long-term effects on evolutionary outcomes. Evolutionary prediction is always challenging, as subtle nuances of dynamics can interact in unpredictable ways. Digital evolution systems, however, provide an empirical alternative to prediction: automated replay experiments can be conducted in large numbers to measure a real distribution of outcomes from a given starting point. Changes in distributions over time can help us understand the long-term implications of seemingly minor events during evolution. We apply this technique to "adaptive momentum", a new framework that explains how phenomena like selective sweeps can temporarily weaken selection and enhance the likelihood of crossing deleterious fitness valleys. We show that deleterious mutations along the leading edge of a selective sweep can have an outsized influence on the evolutionary fate of a population. Indeed, we see that evolutionary potential to cross new deleterious valleys drastically increases during selective sweeps. Moreover, each valley crossing initiates a new sweep, increasing the potential for further discoveries; this increased potential subsides only once all sweeps have concluded. While we still have much to learn about both adaptive momentum and the role of history in evolution, this work identifies important evolutionary dynamics at play and hones our tools for further studies.

Introduction

Innovations in science and technology periodically create opportunities to conduct experimental studies that were previously relegated to the realm of thought experiments. This shift occurred with Stephen Jay Gould's idea of "replaying the tape of life" – starting evolution over again to see if we would arrive at similar outcomes (Gould, 1990). Previous researchers have brought this thought experiment into reality by leveraging microbial populations that can be frozen and then revived (Blount et al., 2018) or digital populations that can be saved and loaded at will (Ferguson and Ofria, 2023). By restarting evolution from different time points in an evolved population's history, researchers have begun to formalize *analytic replay experiments* to test hypotheses on historical contingency. Below we explore how these techniques can be expanded to develop a deeper understanding of evolutionary dynamics, in this case exploring the concept of "adaptive momentum".

Replay Experiments and Evolutionary Prediction

Traditional evolutionary biology is often focused on improving our ability to predict evolutionary outcomes or understand why evolutionary history played out the way it did. Prediction can be especially difficult in light of complex fitness landscapes with epistatic interactions, meaning that the result of mutational combinations cannot always be known based on individual effects. Given the speed of digital evolution, however, we do not always need to predict evolutionary trajectories; we can often empirically measure the range and distribution of possible outcomes. We can even directly measure historical contingency by conducting replay experiments before and after a particular event (such as one or more mutations).

Thus far, these analytic replay experiments have been employed to study the genetic potentiation of complex traits, such as citrate metabolism in *E. coli* (Blount et al., 2008), novel receptor usage in Phage λ (Meyer et al., 2012), and associative learning in digital organisms (Ferguson and Ofria, 2023). Instead of focusing on the evolution of a particular target trait, here we apply replay experiments to an idealized model system to study fundamental evolutionary phenomena.

Adaptive Momentum

The adaptive momentum framework suggests that periods of disequilibrium resulting from phenomena like selective sweeps and range expansions can enhance genetic exploration (Bohm et al., 2024). We use analytic replay experiments to investigate this effect in selective sweeps. Consider the appearance of a beneficial mutation in an asexual spatial population. If this mutation establishes and is sufficiently strong, it can trigger a selective sweep where the genotypes with the beneficial mutation come to dominate the population. During the sweep, there will be a boundary between individuals with the beneficial mutation and those without (wild type). If advantaged individuals along this boundary accrue relatively small deleterious mutations, they may still

have a combined fitness benefit over the wild type. Thus, individuals along the leading edge of the sweep will have an increased potential to accumulate deleterious mutations that, in turn, increase the potential to explore genetic space and facilitate genetic discovery across fitness valleys. Adaptive momentum persists until the wild type is eradicated and equilibrium is reestablished. However, if during fixation a new beneficial mutation is discovered, the state of disequilibrium will persist, thus extending the "momentum window" (the period of adaptive momentum). [1]

In the work presented here, we create an ideal environment for adaptive momentum by using a one-dimensional spatial population evolving on a rising sawtooth fitness function. We use replay experiments to measure how the state of the population affects the potential to cross the next fitness valley. Replay experiments show that adaptive momentum increases the potential for populations to cross fitness valleys, an effect that diminishes over time.

Using simple assumptions about the structure of the leading edge of a selective sweep, we generated a predictive model of the potential for valley crossing during adaptive momentum. While this estimation is highly accurate for early steps into the fitness valley, it loses accuracy when the leading edge is deeper in the valley. The time required for a population to reach deeper mutations in the fitness valley is likely also increasing the number of stochastic events that cause differences from the assumptions of the model. Finally, by shuffling organism positions in our population snapshots we can disrupt population structure. This shuffled analysis allowed us to verify that the organization of the population, not just its genetic composition, is vital to valley-crossing potential. Overall, this work refines the framework proposed by adaptive momentum while advancing the methodology of analytic replay experiments as a tool for studying historical contingency, exposing sections of both that are ripe for further study.

Methods

Evolution system

In order to isolate the adaptive momentum effect and keep computational costs feasible, we used a minimal agent-based evolution model where each organism's genome consists of a single integer, with value x. Fitness is based in a repeated "sawtooth" function, similar to the one used in (Bohm et al., 2024), to create a landscape with regular fitness peaks separated by valleys. Each organism is assigned a quality score ($s(x)$), and then that score exponentiated to determine fitness ($f(x) = 10^{s(x)}$). Score is calculated using the sawtooth function:

$$s(x) = \frac{x - (x \bmod w)}{w} G - (x \bmod w) D$$

where G is the gain per peak, D is the decrease per step into the valley, and w is the valley width (number of mutational steps from one peak to the next). Effectively, the score function calculates the quality of the highest peak achieved and subtracts the cost of the current mutational step into the valley. We used $G = 1.0$, $D = 0.05$, and $w = 6$. This means that peaks appear every six steps; in this case at $x = 6$, $x = 12$, $x = 18$, and so forth. We refer to these peaks by their height, so those three peaks would be p_1, p_2, and p_3, respectively. We refer to the steps between peaks by the prior peak and an offset (*e.g.*, $x = 17$ is $p_2 + 5$, the last step before p_3). Figure 1 illustrates this sawtooth function and the relevant peaks.

Crossing a valley increases an organism's fitness by a factor of 10, and each step into the valley reduces fitness by ~12%, relative to the previous peak. We selected these drastic fitness differences to clearly demonstrate the effect.While this degree of selective benefit may be rare in biological systems, cases such as antibiotic resistance can show fitness increases of this magnitude (Gullberg et al., 2011).

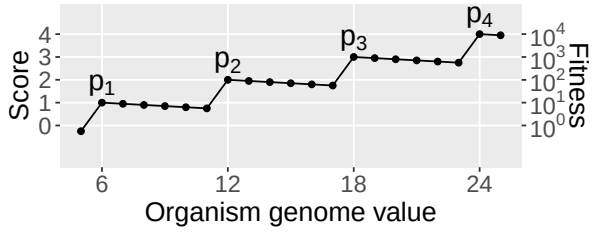

Figure 1: The sawtooth function used in this work, with the four peaks mentioned in this work labeled. Score, $s(x)$, and fitness are both shown, with fitness being $10^{s(x)}$.

We evolved populations of 512 organisms in a one-dimensional spatial population – a single line of organisms – that did not wrap. This structure maximizes fixation times and makes selective sweeps easy to track, as they could only move left or right. The population evolved via discrete, non-overlapping generations. To fill positions in the next generation, we performed rounds of spatial roulette selection, each involving three organisms: the organism in that position in the previous generation and its two immediate neighbors (unless the organism is on the edge, in which case the roulette is between only two organisms). We then copied the selected individual to produce the offspring, with a 0.0125 chance to mutate; mutations either increment or decrement the offspring's genome value by one. Selective sweeps are

[1] Adaptive momentum describes how disequilibrium during evolution can result in periods of increased mutation buffering. Such disequilibrium can be caused by several conditions, including sweeps, range expansions, and increases in carrying capacity in both spatial and well-mixed populations. The adaptive momentum framework further considers how the increased potential for mutational buffering can also affect large scale evolutionary rates via increased genetic exploration and subsequent genetic discovery.

thus limited to advancing one position in the population per generation, limiting the growth rate and therefore the fixation time of a beneficial mutation.

Experiment design

We conducted this work in three stages: 1) we validated that our model demonstrates adaptive momentum; 2) we ran "benchmarking" data to create an expectation of how potentiation changes during a momentum window; and 3) we conducted analytic replay experiments to observe changes in potentiation in evolved lineages. Here we outline these experiments in more detail.

Experiment I: Model Validation To ensure that our model could produce the adaptive momentum effect, we replicated the primary experiment from (Bohm et al., 2024) comparing crossing times in populations starting from equilibrium versus those starting during the disequilibrium of a previous selective sweep. We ran 500,000 evolutionary replicates, where each replicate started with 512 organisms at p_1 and ran for 5,000 generations. In replicates where p_2 was discovered, we recorded the generation of discovery and extended the run duration for another 5,000 generations beyond the point of discovery. If p_3 was discovered before time ran out, we recorded this time as well. This methodology produced two time distributions: time to first crossing and time between first and second crossing, both capped at 5,000 generations.

Experiment II: Benchmarking The adaptive momentum framework posits that populations in disequilibrium experience an increased rate of adaptation. In spatial populations experiencing a selective sweep, the disequilibrium should manifest near the leading edge of the sweep. Specifically, adaptive momentum allows deleterious mutations to accumulate within the advantaged subpopulation along the leading edge. This temporarily expanded mutant cloud increases genetic exploration, accounting for an observed increase in the rate of adaptation.

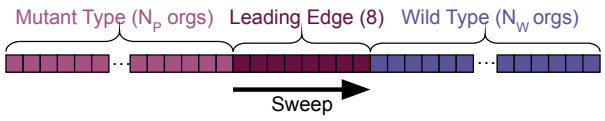

Figure 2: Starting layout for experiment II populations, with three clearly-defined sections: N_P "post-sweep" mutant organisms at peak p_2 (left), N_W "wild type" organisms at peak p_1 (right), and 8 "leading edge" organisms at a treatment-specific position in the valley past p_2 (middle).

We created idealized scenarios to study the dynamics of adaptive momentum as they unfold. Each population had organisms on p_2 sweeping across organisms on p_1, with a well-defined leading edge (Fig. 2). Experimental treatments used all combinations of how far into the next fitness valley the leading edge started (from $x = p_2$ to $x = p_2 + 5$), and how far the sweep had progressed across the population (from $N_P = 0$, at the beginning of a sweep, to $N_P = 504$ at the end.) We ran each replicate for $768 - N_P$ total generations; the reduced number for larger N_P was used to make the comparison fair, subtracting off the minimum time it could have taken to establish N_P post-sweep organisms. For each condition, we measured how often replicates crossed the next valley to reach p_3.

We recorded the number of replicates that successfully crossed to p_3 in each treatment. This measurement provided an expectation of a population's potential to cross the valley *based only on the initial state of the leading edge*. Additionally, we also ran "shuffled" controls under otherwise identical conditions, but where we removed population structure (and thus the notion of a leading edge) by randomly shuffling the organisms before each evolutionary replicate.

Experiment III: Analytic replays of evolved lineages Finally, we focused on the treatment from Experiment II that represented the start of a selective sweep; that is, no post-sweep organisms ($N_P = 0$) and a leading edge that just made it to p_2. We ran 500 replicates under these conditions, saving snapshots at each generation, allowing us to perfectly recreate the population at any time point. We randomly selected 10 replicates that failed to reach p_3, 10 random replicates that did reach p_3 (but not further), and all four replicates that crossed two valleys to reach p_4. For each of these 24 replicates, we performed 1000 analytic replays at every fourth generation, restarting evolution from a given time point with different random seeds to investigate the role of chance in determining *the distribution of potential evolutionary outcomes* (Blount et al., 2018). Next, we selected one representative sample from each of the three categories to study at high resolution, replaying 10,000 replicates from every generation.

Using these replays, we recorded the probability that a replicate would cross the valley to p_3 or p_4 at each time point. These data allow us to calculate how the crossing probabilities changed over time. We also ran 10,000 equilibrium replicates and replayed representative replicates in the same manner. Finally, just as in the shuffle benchmarking experiment, we performed "shuffled" replays on disequilibrium replicates that crossed a single valley. Specifically, we shuffled the population before starting each replay to measure the role of the population's spatial organization in crossing potential.

Data and software availability

All code, analyses, summarized data, and figures not included in this work are available in the supplemental material (Ferguson, 2024). The model was built using the Modular Agent Based Evolver version 2.0 (MABE2) (`https://`

`github.com/mercere99/MABE2`). All analyses and plots were generated using the R statistical computing language version 4.1.2 (R Core Team, 2021) and the ggplot2 (Wickham et al., 2020), dplyr (Wickham et al., 2022), HMisc (Harrell Jr, 2023), tidyr (Wickham and Girlich, 2022) packages.

Results

Validation of adaptive momentum effect

First, we measured the time it took for replicates starting from a full population of p_2 to cross the valley to p_3 and, when relevant, from p_3 to p_4. Figure 3A shows the timing distributions of populations that crossed the first valley over the first 5,000 generations (purple) as well as the timing distributions of second crossings that occur within 5,000 generations of a first (orange). We see that the time of first crossing appear uniformly distributed across the 5,000 generations, while the second crosses are strongly skewed toward shorter time periods. Indeed, of 500,000 replicates, 6,485 crossed the first valley within 5,000 generations (~1.3%) with a mean cross time of ~2,569 generations and a median of 2,602 generations. Based on these values, we would expect roughly 84 replicates to cross twice ($6,485 \times 1.3\%$, or 0.0169% of all 500,000 replicates), but instead we see 902 replicates (~0.18%) cross the second valley, a much higher rate than expected if the probabilities of first and second crossing were equal. In addition to a higher than expected rate of crossing, the mean cross time between first and second crossings is ~579 generations and the median cross time is 401 generations, substantially lower than the first crossing times. Finally, when we consider only those second crossing times that occurred more than 1000 generations after a first crossing, we find that the rate of these second crossings is similar to the first crossing rate (~1.23%). These results comport with the framework of adaptive momentum. They show that an initial beneficial discovery can quickly lead to additional discoveries (during the fixation period), but if the second discovery does not happen before equilibrium is reestablished, the rate of valley crossing is better predicted by the first valley crossing times.

Empirical benchmarks

The empirical benchmark data (Fig. 3B) illustrate how the initial state of the population affects the potential to cross valleys. As expected, steps further into the valley increase the probability of crossing, regardless of where the leading edge is. On the other hand, the probability of crossing decreases as the ratio of p_2 organisms (mutant type) increases relative to p_1 organisms (wild type) – as the selective sweep progresses, fewer opportunities remain for additional mutations to accumulate. While the potential to cross varies considerably with the type of organisms in the leading edge, these data clearly show that all conditions describing early sweep conditions (*i.e.*, having a leading edge and a significant ratio of the remaining population on p_1) substantially increase the probability of crossing the valley compared to populations that are close to reaching equilibrium on p_2.

We use these results to provide a baseline prediction for the analytic replay experiments.

Analytic replay experiments

Of our 500 initial replicates to find candidates for replay experiments, four replicates (0.8%) crossed two valleys in the allotted time, 83 crossed exactly one valley (16.6%), and the remaining 413 did not cross any valleys (82.6%). We randomly-sampled 10 no-cross replicates, 10 single-cross replicates, and took all four double-cross replicates to run coarse-grained replays. All plots are available in the supplement (Ferguson, 2024).

From these coarse-grained replays, we selected one representative replicate from each category to re-run, conducting 10,000 replay trials at *every* time point to give a fine-grained view. For each replayed replicate, we show the potential of crossing the next valley over time, paired with a Muller plot (Muller, 1932) of the initial replicate that was replayed. Figures 4, 5, and 6 show the single-cross, double-cross, and no-cross replicates, respectively. In each plot, the replay results are overlaid on an image generated from the benchmark data. To generate the background image, we treat the benchmark data as a lookup function. When we start a replay replicate, we initialize the population using the snapshot from the target generation of the initial replicate. These snapshots show that the leading edge does not perfectly advance one position every generation; there are many generations where the leading edge either fails to advance or is pushed back one position by the wild type organisms. To make this comparison fair, we find the leading edge in that particular snapshot and then look up the corresponding expectation values from the benchmark data. This adjustment ensures that we are comparing against the correct benchmark data regardless of the motion of the leading edge.

Overall, the potentiation observed in all three replicates closely match the benchmark expectations. The potentiation occasionally increases or decreases suddenly; tracing these changes to the Muller plots typically shows that these events correlate to the gain or loss of a mutation at (or near) the leading edge at that time. For example, the two temporary peaks in potential in Figure 4, at roughly generations 250 and 375, can be directly traced to the leading edge temporarily dipping to $p_2 + 4$. The potentiation at any particular step in the valley decreases over time, as the selective sweep progresses and the adaptive momentum window closes. Figure 6 shows that, while the replicate had substantial potentiation at times (briefly above 50%), it failed to capitalize before the window closed. This failure was exacerbated by two leading mutations from p_2+3 to p_2+2, the second of which dropped the potential from over 25% to under 10%, after which the

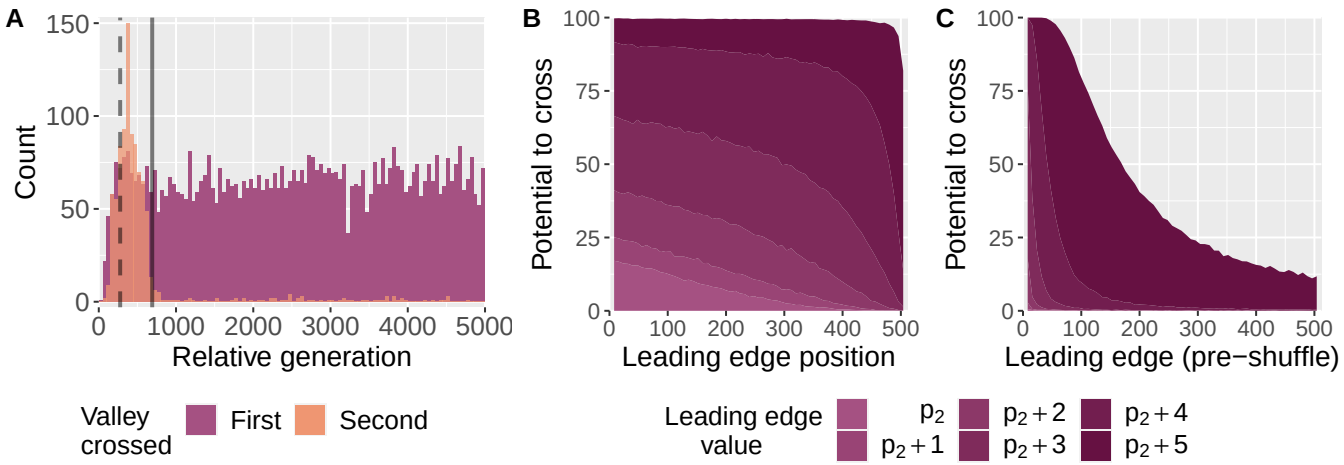

Figure 3: **(A)** Distribution of the number of generations required to cross valleys in the validation experiment. Relative generations refer to the elapsed time (in generations) for the first cross (purple), and from the first cross to the second cross (orange). The dashed and solid vertical lines show minimum and maximum fixation times, respectively. **(B)** Benchmarking data showing the potential of a leading edge to cross a valley, with a range of leading edge starting positions and values (see Fig. 2). Each point represents 10,000 replicates. **(C)** Shuffled benchmarking data. Each replicate is shuffled prior to evolution, otherwise identical to center plot.

population never recovers.

All three replicates show periods of potentiation higher than what our benchmark data would predict given their leading edge genomic value. The benchmarking data modeled a leading edge of eight organisms, but the Muller plots show that the leading edge grows and shrinks over time. The Muller plots show that underprediction is generally associated with either an expanded leading edge or an excess of individuals with lower fitness behind the leading edge. We hypothesize that these underpredictions are generally observed in deeper steps into the valley because of the historical contingency required to reach that point (*i.e.*, to reach $p_2 + 5$ the leading edge must have passed through $p_2 + 4$, which may still exist behind the leading edge).

For comparison, we also ran 10,000 replicates that started at equilibrium (*i.e.*, not in a momentum window). Of those, we replayed the one replicate that crossed twice, 10 randomly sampled replicates from 19 that crossed once, and 10 random replicates from the 9,980 that failed to cross. Figure 7 shows the replicate that crossed twice. The first crossing in this replicate occurred quite early, crossing the valley on generation 168. However, the potential before crossing is similar to all first-cross replicates: while the potential in an adaptive momentum window starts at ~17%, the potential for replicates outside momentum windows starts at ~0.2%. This low potential continues for the first 143 generations, followed by 16 generations with an average potential of ~1.8%, four generations between 15% and 20%, and then the cross. This pure chance, "all or nothing" potentiation is not unique to this replicate; the same dynamic can be seen in the ten other successful replicates we analyzed (Ferguson, 2024).

Finally, we also show the potentiation of crossing the second valley, which in some cases is realized (Figs. 5 and 7), and other times is not (Figs. 4 and 6). As potentiation is probabilistic in nature, the potential to cross the second valley is the probability of crossing the first valley from the current state of the population times the probability of crossing a second valley from a naïve starting position (*i.e.*, the elevated potentiation of a population at the beginning of a sweep). We see this dynamic early in the replays – increases in first-cross potential are reflected, at smaller scales, in second-cross potential, with a much larger increase upon successful crossing of the first valley in Figures 4 and 5. This result is consistent with the adaptive momentum framework, which posits that the discovery of a new peak will initiate a new adaptive momentum window. Strikingly, this same dynamic holds true for Figure 7 – while the replicate did not start in a momentum window, the first cross creates a window which increases the potential of the second cross.

Shuffled population experiments

To test the importance of population structure for potentiation, we repeated the benchmarking and replay experiments with shuffled populations. In both cases, we kept the experiments identical except for an additional shuffle step: before starting a replicate, we shuffled the order of organisms in the population and then proceeded with evolution as normal.

The shuffled benchmark data in Figure 3C shows that only populations with a leading edge of $p_2 + 5$ are able to main-

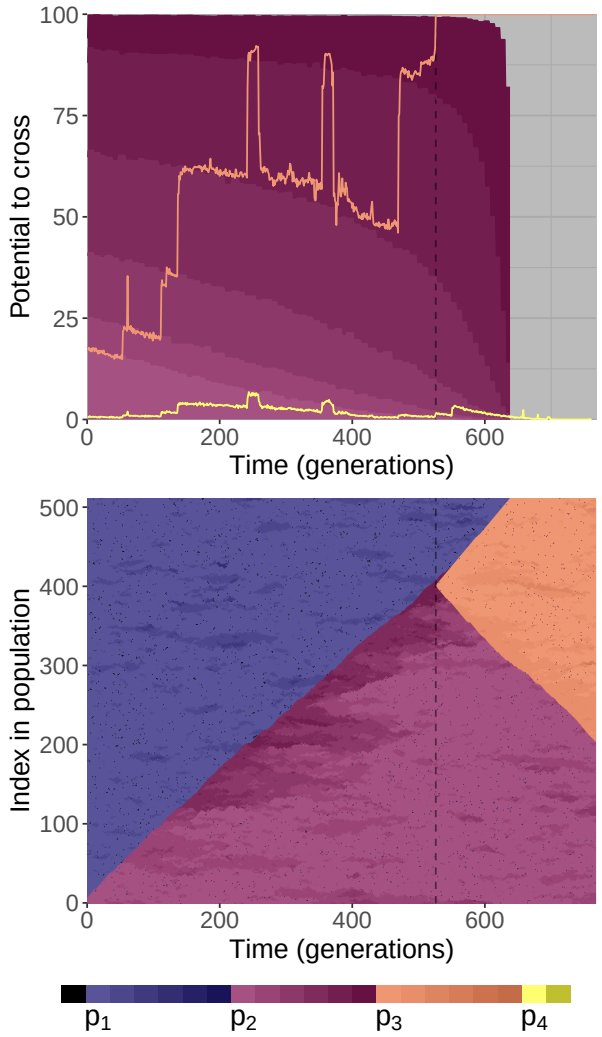

Figure 4: **Top plot**: Analytic replay data for a representative replicate that crossed one valley during a momentum window, overlayed on baseline data. Lines show the probability of crossing both the first valley (orange line; initially-higher) and the second valley (yellow line; initially-lower). The background shows the expected potential to cross (as shown in Fig. 3B); data here are shifted to align with the realized leading edge position. **Bottom plot**: A Muller plot of all organisms in the original 1D population over time. Dark colors show descent into valleys, with hue identifying the valley being crossed. In both plots, vertical dashed lines show initial valley crosses. Colors in the legend apply to all plots.

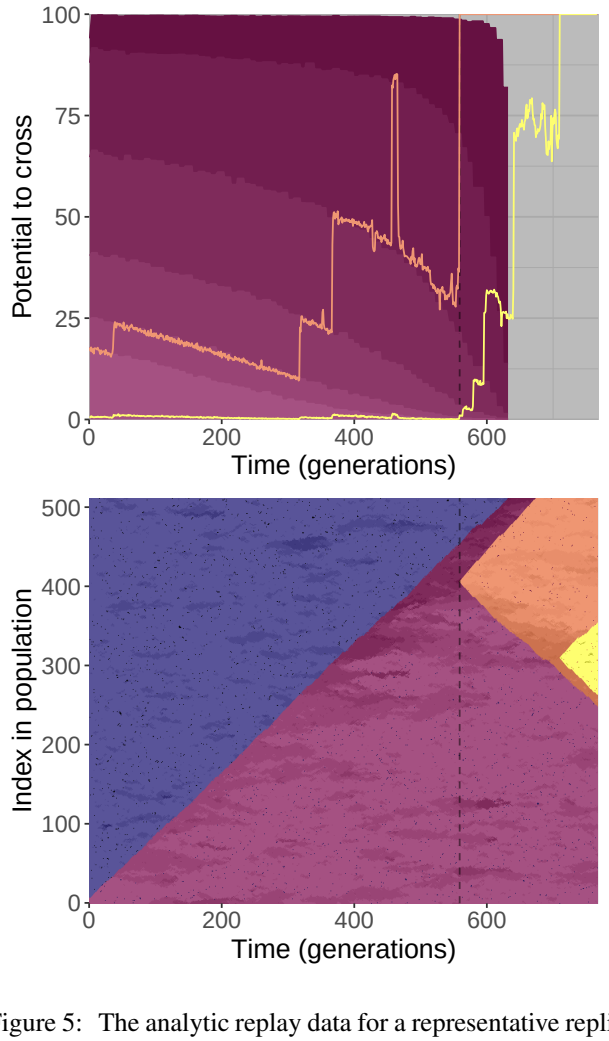

Figure 5: The analytic replay data for a representative replicate that crossed two valleys during a momentum window. See Figure 4 for details and legend.

tain greater than a 10% chance to cross if the leading edge has swept more than half the population. Two differences can decrease crossing potential: 1) multiple leading edges can form, allowing faster fixation and 2) the eight leading-edge organisms are more likely to encounter higher-fitness mutant-type organisms and thus be purified faster.

A representative sample from the single-cross replicates is shown in Figure 8. The replay data indicate that potential to cross remains relatively low, never reaching a 20% chance, until a critical mass of nearly-crossed organisms skyrockets the potential from 9.2% to 100%. The earlier spikes in potentiation always correspond to the appearance of a single organism that is one step away from crossing the valley, which was then lost in the original population. These trends are consistent across all 10 replicates that were replayed (Ferguson, 2024).

Discussion and Conclusion

Potentiation exhibits adaptive momentum

In this work, we have corroborated adaptive momentum's benefit to valley crossing (as outlined in (Bohm et al., 2024))

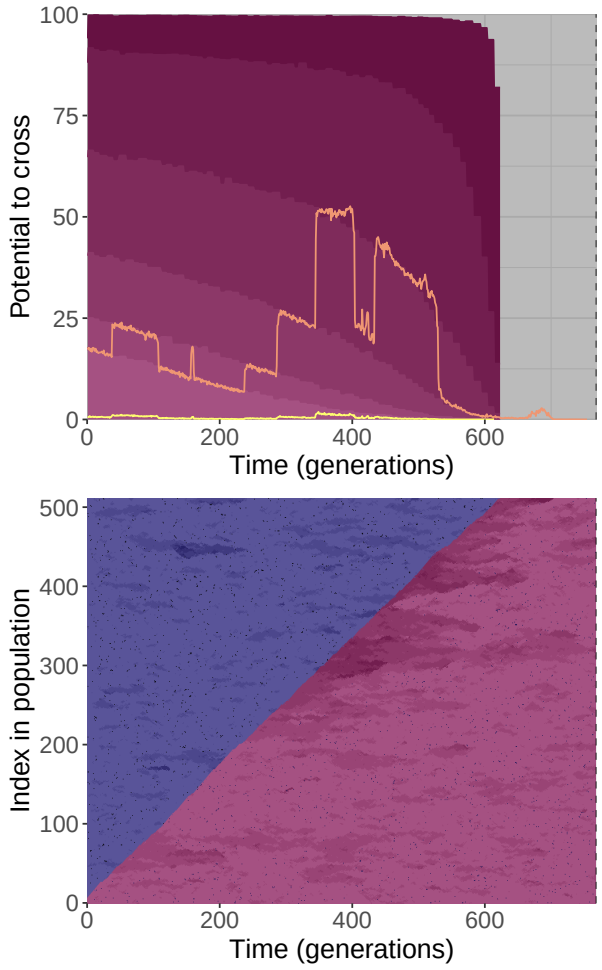

Figure 6: The analytic replay data for a representative replicate that failed to cross a valley during a momentum window. See Figure 4 for details and legend.

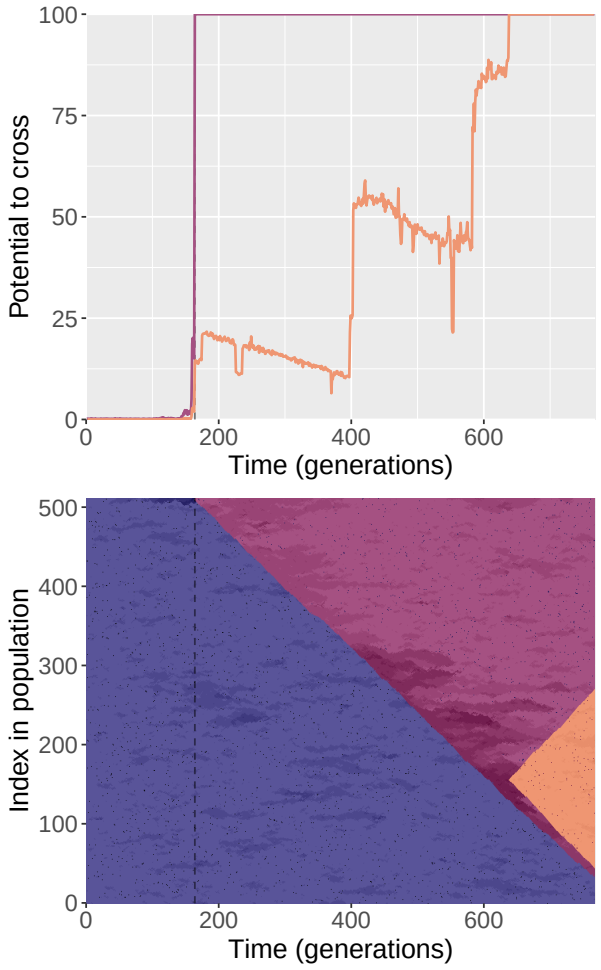

Figure 7: **Top plot**: The analytic replay data for the sole replicate that did *not* start in a momentum window, but still managed to cross a valley, and indeed crossed twice. The red line shows the potential to cross the first valley, while the orange line shows the potential to cross the second valley, exhibiting the hallmarks of adaptive momentum due to the first cross. **Bottom plot**: a Muller plot of the original population, like in Figure 4.

and expanded our understanding of the dynamic. Our initial experiments demonstrated that our system can undergo adaptive momentum, and that disequilibrium is the key driver. However, our main goal in this paper is to provide an alternate vantage point from which to view the dynamics of adaptive momentum. While the original paper validated the effect via aggregated data, here we analyze the underlying dynamics in action on individual populations by quantifying potentiation via replay experiments.

We have shown that populations outside of momentum windows must rely on chance alone to cross a valley. During these "equilibrium" periods, early mutations into the valley, although required for a valley crossing, have no substantial impact on the population's probability of crossing. Conversely, populations in momentum windows immediately see a drastically higher chance to cross, and every mutation in the leading edge is either potentiating or anti-potentiating, depending on the direction. This result is highlighted in Figure 7, where the first cross was pure chance while the second cross was driven by adaptive momentum. This work is only a beginning; future work should apply these techniques to examine the role of adaptive momentum in more complex and realistic fitness landscapes.

The leading edge in a spatial selective sweep determines potentiation

The adaptive momentum framework posits that disequilibrium in a population can reduce the selection pressure on the advantaged subpopulation (Bohm et al., 2024). In spatial populations, disequilibrium is focused at the leading edge

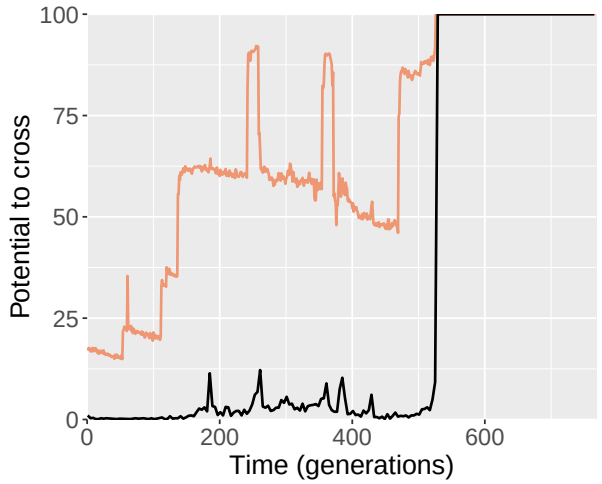

Figure 8: The black line (bottom) shows the potential for a shuffled population to cross a valley. The orange line (top) shows the standard potential for the population to cross (same data as Fig. 4). The shuffled line consists of 1,000 samples every 4 generations.

of selective sweeps, where advantaged mutants encroach on the wild type. Indeed, we observe that the genotype of the organism at the leading edge of the sweep is a strong predictor of potentiation. Moreover, as we can see in the Muller plots, there are often mutated organisms lagging the leading edge. When the leading edge is further into the valley, those lagging organisms have a non-negligible chance to accumulate sufficient mutations to finish crossing the valley, thus increasing the potential.

We aimed to create the simplest system for studying valley crossing in spatial populations. In these one-dimensional populations, our artificially-started sweeps have exactly one leading edge. These edges become more complex in two-dimensional digital systems and only increase in complexity moving toward more natural systems. While the identification and measurements of the leading edge may become more difficult, we expect similar dynamics to hold. It is critical that we continue to improve our understanding of adaptive momentum in these simple systems to build a solid theoretical foundation.

Population heterogeneity and structure affect potentiation

Previous experiments have conducted analytic replays starting from clonal populations (Blount et al., 2008; Ferguson and Ofria, 2023). Here, however, we used perfect population snapshots that record every organism in every generation. Our technique provides a more fine-grained look into how potentiation changes. Furthermore, previous work has often used the most abundant genotype at a time point to seed replay experiments. Across all of our experiments, the most abundant genotype was always on one of the peaks. Replays looking at the potential of the first cross would see effectively zero probability for p_1 and p_2 and 100% probability for p_3 and p_4, which have already crossed, missing all nuance and change in this potential.

The nuance gained by replaying population snapshots provides important insight into using analytic replay experiments, and puts the large jumps in potentiation seen in previous work into question (Ferguson and Ofria, 2023). While those large jumps in potentiation are valid, they are likely missing important intermediate genotypes or population dynamics. In effect, they show genetic effects isolated from effects of population structure. Future work looking to leverage replay experiments should be aware that population structure and composition can affect evolutionary outcomes, and should thus carefully consider how replays are initialized.

Further, by shuffling the population snapshots we have shown that it is not just the portion of the population with each genotype that matters, but also the structural relationships and interactions among the organisms (Figure 8). Other computational studies will likely be needed to tease apart when and how this organization matters as perfectly preserving structure and composition of natural populations for so many replicates is currently impossible. As such, we should leverage computational models to develop techniques possible in both digital and natural populations, possibly finding a middle ground between a single clonal sample and full population snapshots.

Outlook

This work not only provides evidence to support our understanding of adaptive momentum, but further clarifies the underlying mechanisms. By quantifying the potentiation of valley crosses and relating this measure to population composition and structure, we have provided insights into the historical contingencies and long-term trajectories of populations experiencing adaptive momentum. Further, we advance the methodology of analytic replay experiments by conducting them at a greater scale than previously seen, demonstrating a new use case, and leveraging perfect population snapshots to seed the replay experiments. These advances combine to show that replay experiments can extend genetic potentiation to include the effects of population dynamics. All together, this work constitutes both a step toward better understanding adaptive momentum and a methodological refinement of replay experiments for understanding historical contingency.

Acknowledgements

We thank the reviewers and the MSU BEACON lab for comments. This work was supported by the U.S. National Science Foundation (DBI-0939454) and compute resources from the MSU Institute for Cyber-Enabled Research.

References

Blount, Z. D., Borland, C. Z., and Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proceedings of the National Academy of Sciences*, 105(23):7899–7906.

Blount, Z. D., Lenski, R. E., and Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science*, 362(6415):eaam5979.

Bohm, C., Ragusa, V. R., Ofria, C., Lenski, R. E., and Adami, C. (2024). Reduced selection during sweeps lead to adaptive momentum on rugged landscapes. *bioRxiv : the preprint server for biology*.

Ferguson, A. J. (2024). ALife 2024 Supplement. Zenodo and Github. https://doi.org/10.5281/zenodo.11507982.

Ferguson, A. J. and Ofria, C. (2023). Potentiating Mutations Facilitate the Evolution of Associative Learning in Digital Organisms. In *ALIFE 2023: Ghost in the Machine: Proceedings of the 2023 Artificial Life Conference*. MIT Press.

Gould, S. J. (1990). *Wonderful Life: The Burgess Shale and the Nature of History*. WW Norton & Company.

Gullberg, E., Cao, S., Berg, O. G., Ilbäck, C., Sandegren, L., Hughes, D., and Andersson, D. I. (2011). Selection of resistant bacteria at very low antibiotic concentrations. *PLoS pathogens*, 7(7):e1002158.

Harrell Jr, F. E. (2023). *Hmisc: Harrell Miscellaneous*.

Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E., Quick, R. T., and Lenski, R. E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*, 335(6067):428–432.

Muller, H. J. (1932). Some Genetic Aspects of Sex. *The American Naturalist*, 66(703):118–138.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., and Dunnington, D. (2020). *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*.

Wickham, H., François, R., Henry, L., and Müller, K. (2022). *Dplyr: A Grammar of Data Manipulation*.

Wickham, H. and Girlich, M. (2022). *Tidyr: Tidy Messy Data*.