



How the footprint of history shapes the evolution of digital organisms

Jason N. Bundy, Charles Ofria, Richard E. Lenski

Table of contents

- 1. Introduction
- 2. Installation
- 3. Usage
- 4. Development

1. Introduction

In order to emphasize the importance of historical contingency in evolution, Stephen Jay Gould famously proposed a thought experiment of “replaying life’s tape.” Replayng evolution has since provided a conceptual framework in experimental evolution, one that allows researchers to measure the relative contributions of adaptation, chance, and phyletic history (Blount, Lenski, and Losos, 2018, *Science*). Several relatively short experiments with microbes have shown that strong selection in a new environment can overwhelm the signature of history (e.g., Travisano et al., 1995, *Science*). However, it has been suggested that over much longer periods in the ancestral environment, the “footprint of history”—the genomic legacy of prior adaptation—might become too deeply engrained to overcome even with intense selection in a new

environment. Addressing this issue experimentally has remained elusive owing to the amount of time required to meaningfully extend the phylogenetic depth of an evolution experiment and the volume of lab work needed to assess the relative contributions of adaptation, chance, and history as a function of multiple phylogenetic depths.

We overcame these obstacles by performing a two-phase evolution experiment with digital organisms using the Avida platform. (Evolution experiments using Avida have been published in PLoS Biology, Science, Nature, and elsewhere.) In the first phase, we evolved 10 populations from a common ancestor under identical conditions for ~60,000 generations. We then sampled clones from each population at three timepoints corresponding to shallow, intermediate, and deep history. In the second phase, we evolved 10 populations from each of these 30 founders for many generations in two distinct new environments, one similar and one dissimilar to the ancestral environment. We then measured the contributions of adaptation, chance, and each founder's history to the evolution of both fitness and genome length. We show that the footprint of history can indeed become too deep to be overwhelmed by selection. Importantly, we also show that this effect depends on the strength of selection acting on the focal trait and the degree of similarity between the ancestral and new environments.

2. Installation

The latest version of Avida can be downloaded at <https://avida.devo.org/>. If you are new to Avida, you can find helpful resources for new users on the Wiki <https://github.com/devosoft/avida/wiki/BEGINNER-Doc>. The instructions provided here assume familiarity with Avida and its basic configuration files. Further, you are free to modify and alter any of the included configuration files for your own projects which may utilize any directory structure that suits you. The instructions provided here will refer to directory structure and configuration files we used in support of the current manuscript.

Directory Structure

In your primary project directory you will want to have the following sub-directories (i.e. folders):

```
-config: The directory contains the Avida's setup and configuration files.  
-logs: This directory is for storing the text files that record information about various scripts.  
-output_phase1: This directory will contain the data output by Avida during the first phase of the experiment.  
-output_phase2: This directory will contain the data output by Avida during the second phase of the experiment.  
-output_analysis: This directory will contain the files and subfolders that result from analyzing the data.  
-python_scripts: This directory contains the python scripts necessary for analyzing the data. It contains 'analysis' and 'operations' sub-directories.  
-resources: This directory contains a subfolder for storing fonts and can be used as a container for additional project resources.
```

config

This folder contains the Avida executable (avida), the Avida configuration file (avida.cfg), the analyze mode configuration file (analyze.cfg), and a subdirectory containing folders for various types of setup files. The

"setup" subdirectory contains folders for environment files, events lists, instruction sets, saved organisms, and saved populations.

logs

This folder contains simple text file that record the operation and output of various python scripts used in the analysis. These are similar to a "lab notebook" generated automatically by the scripts. They can be useful in keeping track of the project as well as having clear textual representations of what the various scripts have done. They can also be useful in detecting errors.

output_analysis

This will contain dataframes and two subdirectories. One, "raw", for copies of unprocessed dataframes used during the analysis and the other, "masked", for filtered datasets, plots, and figures that result from running various analysis scripts.

output_phase1

This will contain the individual output directories that correspond to each individual run. Each individual output directory will contain a "data" subfolder with the individual run's data, a "setup" subfolder for storing initial setup files (i.e. environments, events, instruction sets, saved organisms, and saved populations), as well as the avida executable file, the analyze mode configuration file (i.e. 'analyze.cfg'), and the Avida configuration file (i.e. 'avida.cfg').

output_phase2

This will contain the individual output directories that correspond to each individual run. Each individual output directory will contain a "data" subfolder with the individual run's data, a "setup" subfolder for storing initial setup files (i.e. environments, events, instruction sets, saved organisms, and saved populations), as well as the avida executable file, the analyze mode configuration file (i.e. 'analyze.cfg'), and the Avida configuration file (i.e. 'avida.cfg').

python_scripts

This contains the Python scripts for the project in two subfolders. The first, "analysis" contains scripts used to analyze data, while the other, "operations", contains scripts used in various procedures throughout the project including the creation and sorting of data dicts and related .csv files, label formatting, and miscellaneous operations.

resources

This folder can be used to store any other files or assets you'd like to keep organized with the project. By default, it contains one subfolder, "fonts", for storing fonts that can be used for plotting.

The full directory structure used for this project is shown below. If you'd like to recreate our experiment without modifying the original code you will want to replicate this structure. If you'd like to use a different directory structure, you can use the structure below as a reference and modify the code to suit your needs. It's important to remember that subdirectories labeled "...individual run directories" will be output by Avida and therefore you do not need to manually create the subdirectories (i.e. data and setup).

Project directory:

- config
 - setup
 - environments
 - events
 - instructions
 - organisms
 - populations
- logs
 - created_events
- output_analysis
 - masked
 - ACH_estimates_shallow
 - ACH_estimates_trajectory_shallow
 - anova_tables
 - estimated_variance
 - figures
 - ACH_images
 - footprint_ACH_estimates_bytrait
 - footprint_ACH_estimates_trajectory_bytrait
 - paired_ttest
 - phase1_plot
 - phase2_bnyenvironment_plots
 - phase2_bytrait_plots
 - raw
- output_phase1
 - ...individual run directories
 - data
 - setup
- output_phase2
 - ...individual run directories
 - data
 - setup
- python_scripts
 - analysis
 - operations
 - axis_labels
- resources
 - fonts

3. Usage

We performed our experiment using a custom run system developed by the Digital Evolution Lab at Michigan State University on the High Performance Computing Center (HPCC) maintained by Michigan State University's Institute for Cyber-Enabled Research (ICER). We have included the run_list files we used to perform the experiment. However, below we describe the parameters necessary for recreating our experiment in the absence of these resources. Using the same initial seed should generate the same results.

We used run_list files to execute our runs using MSU's High Performance Computing Cluster. We have included those files (run_list_phase1, run_list_phase2) in this repository for reference. We have also included the configuration files necessary to reproduce our experimental design.

Phase 1

In the first phase of the experiment we evolved ten isolated, replicate populations from a single ancestral genotype in a common environment. You can reproduce our results using the following configuration details with the files included in this repository.

Seed: 101-110 avida configuration file: avida.cfg environment: env_38tasks_noequ_even.cfg events list: events_500k.cfg instruction set: instset-heads.cfg default organism: default_heads.org

Phase 2

In the second phase of the experiment we selected the most dominant genotype from each lineage we evolved during the first phase at three different time points corresponding to a shallow (20k), intermediate (100k), and deep (500k) footprint of history. These genotypes will be saved by Avida in the "archive" folder within the "data" directory for each run. You can identify the appropriate ancestor by searching the .org files to find the files that match the appropriate "Update Output" (i.e. 20000, 100000, or 500000). We then evolved 10 isolated, replicate populations from each of these genotypes for 100000 updates in two new environments, one similar ("Overlapping") and one dissimilar ("Orthogonal") to the environment used for the first phase. You can recreate the second phase of our experimental design by using each selected genotype to initiate 10 runs in each environment with the following configuration details.

Seed: 1001-1010 avida configuration file: avida.cfg environments: "Overlapping"- env_76tasks_noequ.cfg and "Orthogonal"- env_38tasks_noequ_odd.cfg events list: events_100k.cfg instruction set: instset-heads.cfg default organism: ... appropriate .org file saved from each lineage from the first phase

Analysis

The primary data for each run following the second phase will be found in the "average.dat" file within the "data" directory for each run. The data we used in our analysis are labeled "fitness" and "copied size". We used a nested variance components analysis for each trait in each environment to estimate the contributions of "chance" (measured as the variation within replicates derived from the same ancestor) and "history" (measured as the variation between groups of replicates from different ancestors). We estimated the contribution of adaptation as the mean difference in trait values between each set of 100 Phase II populations and their corresponding ancestor. We also plotted these estimates over time throughout Phase II. Please see the manuscript for additional details on our analysis.

4. Development

You may use our configuration files in any way that suits you. These files are provided as is. At present, inquiries and support will only be addressed in support of the present manuscript. To inquire, please email author Jason Bundy (bundyjay86@gmail.com).