

# How the footprint of history shapes the evolution of digital organisms

---

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

## Table of contents

1. Abstract
2. Installation
3. Usage
4. Development

### 1. Abstract

Gould's thought experiment of "replaying life's tape" provides a conceptual framework for experiments that quantify the contributions of adaptation, chance, and history to evolutionary outcomes. Using this framework, one can measure how varying the depth of history in one environment influences subsequent evolution in a new environment. Can the "footprint of history"—the genomic legacy of prior adaptation—grow too deep to overcome? Can this legacy constrain adaptation, even with intense selection in the new environment? We investigated these questions using digital organisms. Specifically, we evolved ten populations from one ancestor under identical conditions. We then replayed evolution from three time points in each population's history (corresponding to shallow, intermediate, and deep history) in two new environments (one similar and one dissimilar to the prior environment). We measured the contributions of adaptation, chance, and history to the among-lineage variation in fitness and genome length in both new environments. In both environments, variation in genome length depended largely on history and chance, not adaptation, indicating weak selection. By contrast, adaptation, chance, and history all contributed to variation in fitness. Whether the depth of history affected adaptation depended crucially on the environment. When the ancestral and new environments favored overlapping traits, history was as important as adaptation to the fitness achieved in the new environment for the populations with the deepest history. However, when the ancestral and new environments favored different traits, adaptation overwhelmed even deep history.

### 2. Installation

The latest version of Avida can be downloaded at <https://avida.devosoft.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-Documentation>. 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 the 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 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 files 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 sub-folders. 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 dictionaries 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\_byenvironment\_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 using 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 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) within each group of lineages for each depth of the footprint of history. We estimated the contribution of adaptation as the mean difference in trait values between each set of 100 populations from each depth of history from the second phase and their corresponding ancestor. We also plotted these estimates over time throughout the second phase. Please see the manuscript for additional details on our analysis.

## 4. Recreating plots and tables

There is a file called "recreating\_visuals.md" in the scripts folder that describes how to create each figure and table in the manuscript.

## 5. Development

Creative Commons CC-BY\_NC: Anyone can share, reuse, remix, or adapt this material, providing this is not done for commercial purposes and the original authors are credited and cited. 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).