

Data archive for:
Genetic regulatory network motifs constrain adaptation through
curvature in the landscape of mutational (co)variance

Tyler D. Hether and Paul A. Hohenlohe

In the sections below we provide information on the 6 directories used for data analysis.

1 Directory `./01-Genotype_to_phenotype_map`

This directory contains the R script ‘Get_Initial_Allelic_Values5.R’ which was used to create contour plots in figure 1 of the main text.

2 Directory `./02-1_generation_6motifs_3x3_analysis`

This directory contains three files and three directories that were used to generate figures 2, S2 and S3.

2.1 Data Preprocessing

Sub-directory `./01-preprocessing` contains the R script `Sim_Infer.r` which is used to define the location of the 9 populations.

2.2 Data Generation

File `RunMe.sh` is a bash script that runs through all the parameter combinations that were defined in the Preprocessing step and allows the user to set the maximum number of cpus to use. File `SimulationCodeCluster.r` generates the data. Sub-directory `./02-runs` (empty) stores the output pdfs (optional). Sub-directory `./03-SummaryOutput` stores the various summary output for each parameter combination.

2.3 Data Analysis

File `Figure02-PostProcessingCode_GridAnalysisFigure2v4.r` takes data that were generated in `./03-SummaryOutput` and plots the mutational (co)variation, additive genetic (co)variation, and epistatic (co)variation).

3 Directory `./03-07-Rescaling_20x20_grid`

This directory contains three files and three sub-directories that were used to generate figures 3, 4, S1, and S4.

3.1 Data Preprocessing

File `Sim_Infer.r` in sub-directory `./01-preprocessing` defines the grid locations.

3.2 Data Generation

File `RunMe.sh` is a bash script that runs through all the parameter combinations (defined in `./01-preprocessing/Sim_Infer.r`) and allows the user to set the maximum number of cpus to use. File `SimulationCodeClusterJustM.r` calculates the M matrix for each position along the 20x20 grid. Sub-directory `./02-runs` (empty) contains the output pdfs (optional). Sub-directory `./03-SummaryOutput` contains the elements of M-matrix for each grid point and for each motif.

3.3 Data Analysis

File `Figure03-07-PostProcessingCode_GridAnalysis6.r` takes the data generated in `./03-SummaryOutput` and plots rescaled phenotypic space for all 6 motifs, rescaled phenotypic space with population means superimposed for motif C, and dimensionality and evolvability of M across trait space for each motif. The data used for plotting mean phenotypes through time are found in directory `./04-1000_gens_6motifs/03-SummaryOutput`.

4 Directory `./04-1000_gens_6motifs`

This directory contains the three files and three sub-directories that were used to generate figure 4 of the main text.

4.1 Data Preprocessing

File `Sim_Infer.r` in sub-directory `./01-preprocessing` defines motif type, number of replicates, selection strength, and migration rate between the two populations.

4.2 Data Generation

File `RunMe.sh` is a bash script that runs through all the parameter combinations (defined in `./01-preprocessing/Sim_Infer.r`) and allows the user to set the maximum number of cpus to use. File `SimulationCodeCluster.r` simulates a two population divergence with gene flow model. Sub-directory `./02-runs` (empty) contains the output pdfs (optional) and sub-directory `./03-SummaryOutput` contains the summary output for each simulation run.

4.3 Data Analysis

File `Figure04-PostProcessingCode20130827_averageFlyingGs.r` takes the data generated in `./03-SummaryOutput` and plots the average (across replicates) G-matrix and average (across replicates) mean phenotypes through time.

5 Directory `./05-06-20000_gens_motifC`

This directory contains the four files and three sub-directories that were use to generate table 1 and figures 5, 6, and S5-S9.

5.1 Data Preprocessing

File `Sim_Infer.r` in sub-directory `./01-preprocessing` defines the number of replicates, selection strength, and migration rate between the two populations.

5.2 Data Generation

File `RunMe.sh` is a bash script that runs through all the parameter combinations (defined in `./01-preprocessing/Sim_Infer.r`) and allows the user to set the maximum number of cpus to use. File `SimulationCodeCluster.r` simulates a two population divergence with gene flow model. Sub-directory `./02-runs` (empty) contains the output pdfs (optional) and sub-directory `./03-SummaryOutput` contains the summary output for each simulation run.

5.3 Data Analysis

File `Figure05-PostProcessingCode20130828_20000gensSummary.r` plots 1) average (across simulated replicates) extent of adaptation for each population for motif C and 2) extent of adaptation individually for each replicate. This file also generates the summary statistics given in table 1 of the main text. Please note when running this script that memory requirements are high because each replicate's summary output contains 20,000 generations of data! If you are working on a computer with limited memory, you might be able to load the smaller saved R image (105.9 MB). To do this, change the name of the file `RData` (found in `./03-SummaryOutput`) to `.RData` (i.e., make it a hidden file so that R will find it upon loading).

File `Figure06-PostProcessingCode20130906_averageFlyingGs_20000gens.r` takes the data generated in `./04-SummaryOutput` and plots the average (across replicates) G-matrix and average (across replicates) mean phenotypes through time.

6 Directory `./Mal_distance`

This directory contains the R script `Mal_distance.R` which was used to generate estimate the 'mutational' and physical (Euclidean) distance traveled for each population and for each motif (averaged over simulated replicates) (table 2). The input data for this code is found in `./04-1000_gens_6motifs/03-SummaryOutput`.