Description of contents

Carlos Espinosa-Soto*, Ulises Hernández, and Yuridia S. Posadas-García

Instituto de Física, Universidad Autónoma de San Luis Potosí, Manuel Nava 6, Zona Universitaria, San Luis Potosí, Mexico

April 3, 2021

Here you can find all the code required to produce the results reported in the main text and supplementary material of the article "Recombination facilitates genetic assimilation of new traits in model gene regulatory networks", authored by us. Unless noticed, the scripts are written in C++. The programs require the GSL numeric library for C++ (`https://www.gnu.org/software/gsl/`). Therefore, compilation using `g++` requires including the flags `-lgsl -lgslcblas`. Our codes save the data that we generate in a specific directory structure that we provide here.

1 libs/

This directory contains libraries with functions that our scripts require. Compilation of most of our code requires linking with some of the libraries in this directory.

2 Sample/

This directory contains the programs required to build large samples of networks. The files in `GraphC` refer to networks in which the weights in the genotype matrix are floating point numbers; files in `GraphI0` concern networks in which the threshold θ_i is always set to 0; files in `GraphI` refer to the standard case, reported in the main text and in most of our analyses.

The `Graph*/` directories contain some of the following directories: i) `sinplas` refers to A networks, in which we do not require to have access through plasticity to a second gene activity pattern (GAP) B; ii) `pn1` refers to networks that produce GAP B after single gene perturbations in the standard initial condition $\mathbf{s}^{\widehat{0}}$; iii) `pn2` contains files to sample networks that produce GAP B after two-gene perturbations in the standard initial condition; iv) `pfp` refers to networks that produce GAP B albeit without any requirement for the initial condition leading to it.

Within the latter subdirectories there are `*.cc` files with the code required to sample network with characteristics corresponding to each subdirectory. The subdirectories also contain `*N.zip` files with the directory structure required to save the networks that we sample. Within this structure, for example, the path `12N/36e/3_3_6`refers to networks with 12 nodes, 36 interactions and a pairwise distance between $\mathbf{s}^{\widehat{0}}$ and A, between $\mathbf{s}^{\widehat{0}}$ and B and between A and B equal to 3, 3 and 6, respectively. At the end of this path there are two files and a directory called `networks`.The file `grabthis.txt`indicates which is the next file of networks that must be moved to incorporate into the analyses. The `*.sh` file is a pbs script that calls the program to sample networks. The `networks/`directory will contain the networks that we sample.

*Corresponding author: carlos.espinosa@uaslp.mx; c.espinosa@ifisica@uaslp.mx

3 CongPlaMu/

This directory contains the programs required to assess the relationship between access to a new phenotype through mutation and through plasticity (without recombination). The files in `GraphC` refer to networks in which the weights in the genotype matrix are floating point numbers; files in `GraphI0` concern networks in which the threshold θ_i is always set to 0; files in `GraphI` refer to the standard case, reported in the main text and in most of our analyses. The file `setupea.cc` contains code to prepare the directory structure for these analyses. The file `jala_redes.cc` provides the code that moves already sampled networks (in the `Sample` directory, described above) into the appropriate location for the analyses. The file `plot.cc` contains the code required to build plots and perform statistics analyses once the results are obtained.

Within the `Graph*` directories there is a file called `CongPlaMu.cc` that contains the code for the analyses. There are also zip files with the appropriate structure to log the data of the analyses of different kinds of networks. As earlier, `sinplas` refers to A networks, in which we do not require to have access through plasticity to a second gene activity pattern (GAP) B; `pn1` refers to networks that produce GAP B after single gene perturbations in the standard initial condition $\widehat{\mathbf{s}^0}$; `pn2` contains files to sample networks that produce GAP B after two-gene perturbations in the standard initial condition; `pfp` refers to networks that produce GAP B albeit without any requirement for the initial condition leading to it. A path, such as `12N/36e/3_3_6` , has the same meaning as described earlier. At the end of this path there are two directories and one file. The file `correle.sh` contains a script that calls the program with code in `CongPlaMu.cc` with appropriate parameters. The directory `networks/` contains the networks moved from the directory `Sample/` and the directory `Results/` will store the results of the analysis.

4 Fam_ABpivot/

This directory contains tools for the comparison between the offspring of $AB{\times}AB$ and $AB{\times}A$ network pairs. The directories `GraphC`, `GraphI` and `GraphI0` distinguish analyses for networks with different properties, as described above. The file `prim_alista.cc` contains code to prepare the directory structure for these analyses. The file `jala_redes.cc` provides the code that moves already sampled networks (in the `Sample` directory) into the appropriate location. The file `plot.cc` contains the code required to build plots and perform statistics analyses once the results are obtained.

Within each `Graph*` directory there is a file `fam_abpiv.cc` with the code for the analyses. In the `GraphI` directory there is an additional file `pagor.cc` that does the same but for the largest networks that we study. The difference is that we do not assess all possible offspring for each pair of such large networks but a random sample of 2^{10} offspring, because of computational cost. The zip files contain the directory structure for the analyses, using the same descriptors (`pn1`, `pn2`and `pfp`) explained above. A path, such as `12N/36e/3_3_6` , has the same meaning as described earlier. At the end of this path there are three directories and a file. The file `*.sh`provides the pbs script to submit jobs with appropriate parameters. The `networks/` directory stores the networks used in the analyses, copied from `Sample/`, and `Results/` will store the output data from the program. The `Figs/` directory will store the results of statistical analyses and output figures.

5 Fam_Apivot/

This directory contains files analogous to those of `Fam_ABpivot/`, described above, but for the comparison between the offspring of $A{\times}A$ and $AB{\times}A$ network pairs.

6 DisInc/

This directory contains files for the analysis of the effect of genetic distance between parents and access to a new phenotype through plasticity, mutation and recombination. The file `prim_alista.cc` contains code to prepare the directory structure for these analyses. The file `jala_redes.cc` provides the code that moves already sampled networks (in the `Sample` directory) into the appropriate location. The file `plot.cc` contains the code required to build plots and perform statistics analyses once the results are obtained. Inside the `GraphI/` directory, the file `increg.cc` contains the code for the analysis. Inside the `pn1.zip` file there is the directory structure required to save the data. A path, such as `12N/36e/3_3_6` , has the same meaning as described earlier. At the end of this path there are two directories and a file. The file is the pbs script to submit jobs to a computer cluster with appropriate parameters. The `networks/`directory stores the networks used in the analyses, copied from `Sample/`, and `Results/`will store the output data from the program.

7 CarPlas/

This directory contains files for the evolutionary simulations designed to address whether ancestral plasticity plays a role in setting the course that evolution takes. Inside the `SexualP1/` directory there are three directories and several files. The file `jala_redes.cc` provides the code that moves already sampled networks (in the `Sample` directory) into the appropriate location. The file `grafica.cc`contains the code required to build plots and perform statistics analyses once the results are obtained. The file `carreras_sex.cc` contains the code required for the simulations. Files `fromA.sh` and `fromAB.sh` are pbs scripts to submit jobs to a computer cluster without and with ancestral plasticity, respectively. The file `nexttowr.txt` logs the number of copied networks from the `Sample` directory. The `networks/` directory stores the networks used in the analyses, copied from `Sample/`, and `Results/` will store the output data from the program. The `Figs/` directory will store the results of statistical analyses and output figures.

8 RA_FGF/

This directory contains the programs required to test our results in a network that was used to model the antagonistic gradients of Retinoic Acid (RA) and Fibroblast Growth Factor (FGF) in the presomitic mesoderm of vertebrates as described in Goldbeter et al. (2007). The file `fa_fgf.m` contains the differential equations used to model the system.

The `Case/` directory contains the following programs: i) `getci.cc` contains code to generate initial concentrations that allows an initial network to produce one specific GAP. Additionally, the network must be able to produce, from a different initial condition, the alternative GAP. ii) `getFGen.m` contains code required to form random networks that produce two GAPs. iii) `access.m` contains code to test the correlation between access to a new GAP through mutation and through plasticity. iv) `fami.m` contains code to analyze the offspring of two networks and test whether offspring of networks that produce an alternative GAP by many different perturbations inherits the ability to produce the alternative GAP as well. v) `reconf.cc`, `access.m.` and `fami.m`contain code to plot and make statistical analysis of the results.

The `libs/` directory contains libraries with general purpose functions that our programs requires. Compilation of `getci.cc` requires linking with the libraries in this directory.

9 toggle/

This directory contains the programs required to test our results in a synthetic network developed by Gardner et al 2000 in *Escherichia coli*. `toggle.m` contains the differential equations used to model the system. The `Case/`directory contains programs analogous to those in `RA_FGF/Case`, but for the model described in Gardner et al 2000.