MetHis documentation

Romain Laurent

December 4, 2020

`MetHis` is a population genetics forward-in-time simulation tool designed to simulate arbitrarily complex admixture histories between two populations.

1 Requirements

Before installing `MetHis`, make sure that the following softwares and libraries are installed on your computer:

- GSL library ($\geq$ 2.1, webpage)

2 Installation

To compile `MetHis`, type "`make`" in the current directory.

3 Parameters generation

We provide the Python script "`generate_params.py`" which can generate parameter files to use in `MetHis` under a wide variety of models. The parameters of this script are:

- `nb_simulation`, the number of parameters sets to generate

- `nb_generation`, the number of generations per simulation

- `prefix`, the name of the directory to store the parameters

- `contrib_s1`, a description of the evolution of the contribution of source population 1 over time (see section 3.1)

- `contrib_s2`, a description of the evolution of the contribution of source population 2 over time (see section 3.1)

- `Ne`, a description of the evolution of the admixted population size over time (see section 3.2)

- `force-rewrite`, a flag (default to False) to overwrite existing files/directories

The script will then create a directory (name defined by `prefix`) containing as many sub-directories as the number of simulation requested (`nb_simulation`) called `simu_1`, ..., `simu_N`. Each sub-directory `simu_i` contains two files:

- `simu_i.txt` contains the values of the parameters of the model generated. It can be used in an ABC framework to perform parameter inference (see section 3.3 for further details)

- `simu_i.par` contains the contributions of each source population and N_e of the admixted population at each generation.

Please note that `MetHis` expects the directories and files to be present and named as described above. Please respect this nomenclature if you use your own parameters generation script.

3.1 Contribution argument

The contribution argument is expressed as follows: $c_0/c_{pattern}/pattern\_arguments$. c_0 is the range of the contribution at foundation of the admixted population, in the form `min-max`, with `min` and `max` between 0 and 1. $c_{pattern}$ is the contribution pattern, one of `Con` (constant), `Dec` (decrease), `Inc` (increase) or `Pulse`. The $pattern\_arguments$ depend on the $c_{pattern}$ choosen:

- with the `Con` pattern, it is the range (in form `min-max`) of the recurring contribution at each generation

- with pattern `Dec` or `Inc`, it is two ranges separated by a "/", representing the range of the contribution at generation 1 and the range of the contribution at the last generation

- with pattern `Pulse`, the first argument is the number N of pulses to generate, followed by N ranges indicating the intensity of each pulse. Everything is separated by "/".

Note that any of the range described above can be replaced by the word "`default`", which is similar to `0.0-1.0`, i.e a completely uninformed prior.

3.2 N_e argument

This argument is of the form: $N_{e_0}/N_{e_{pattern}}/N_{e_{range}}$. N_{e_0} is a single integer representing the admixed population size at foundation. $N_{e_{pattern}}$ is one of `Con`, `Dec` or `Inc`. $N_{e_{range}}$ is a range indicating the admixed population size at the last generation (or at every generation if $N_{e_{pattern}}$ is `Con`). This range can be replaced by "`default`", which is equivalent to `10-5000`.

3.3 Format of the `txt` files generated

Each sub-directory `simu_i` generated contains a `simu_i.txt` file. This file contains the values of the parameters used to generate the corresponding `simu_i.par` file. These parameters are given separately in order to simplify the integration of the simulations in an ABC framework. The parameters given in this file are:

Ne.0 admixed population size at foundation

Ne.1 admixed population size at the first generation after foundation

Ne.N admixed population size at the last generation

Ne.u shape parameter of admixed population size increase/decrease

s1.0 contribution of source population 1 at foundation

s1.1 contribution of source population 1 at the first generation after foundation

s1.N contribution of source population 1 at the last generation

s1.u shape parameter of contribution of source population 1 increase/decrease

t1.X time of admixture pulse X from source population 1

i1.X intensity of admixture pulse X from source population 1

All the parameters regarding source population 1 have their counterpart for source population 2. It should be noted that depending on the scenario simulated, some of the parameters above could be irrelevant, in which case they will not be present in the `txt` file, or their value will be set to `NA`.

4   Running `MetHis`

`MetHis` arguments are explained by running "`./Methis`" with no argument. The `prefix` argument is typically the same as the `prefix` argument for the parameters generation script. The names of the sub-directories and files in this directory are described in section 3. The input genotypes file for source populations is expected to be in Arlequin format, such as the output of `fastsimcoal2` (webpage). Please note that the "`R_codes`" directory should be placed in the same directory as the `MetHis` executable.

After a successfull simulation, the `simu_i` folder will contain a "`final_sumstats.txt`" file containing the summary statistics computed for this simulation. These files can be used in an ABC framework to perform model choice and parameter inference.