**Manual**

**Bait-ER**

a Bayesian method to detect targets of selection in

Evolve-and-Resequence experiments

**Version**

0.0 (December 2019)

**Citation**

soonish

**Questions and bug reporting**

You can contact Rui Borges ([ruiborges23@gmail.com](mailto:ruiborges23@gmail.com)) or Carolina Barata ([cdcbrb@st-andrews.ac.uk](mailto:cdcbrb@st-andrews.ac.uk)) if you have any questions regarding the model/code/output.

**License**

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software. See the GNU General Public License (http://www.gnu.org/licenses/) for more details.

**Download and compile Bait-ER on Linux**

First of all, we need to download all the necessary files to compile and run Bait-ER. These are in our GitHub branch at <https://github.com/mrborges23/Bait-ER>. You can download these files manually of by pulling it directly from GitHub using the terminal:

git clone https://github.com/mrborges23/Bait-ER.git Bait-ER

Before compiling Bait-ER we need to make sure that some packages are installed in your machine. Bait-ER uses armadillo, which requires packages like LAPACK, Boost and BLAS. In Ubuntu, you can simply run the following commands.

sudo apt-get install liblapack-dev

sudo apt-get install libblas-dev

sudo apt-get install libboost-dev

sudo apt-get install libarmadillo-dev

You can the type the following command to compile Bait-ER:

g++ main.cpp -o baiter -O2 -std=c++14 -llapack -lblas -larmadillo

You may get the error that some libraries (e.g. libhdf5.so) could not be found by the compiler. You can solve this problem by finding the location of the file missing and adding its path to the $LD\_LIBRARY\_PATH path. Some example:

locate libhdf5.so

export LD\_LIBRARY\_PATH=$LD\_LIBRARY\_PATH:path/to/missing/file

If the compilation proceeds without errors, you are now able to run Bait-ER.

**Input file**

The input file of Bait-ER is a sync file. It may include an header line, but this is options. A example for the first five loci in the example.txt sync file are:~

The order of the columns can also change; the control file permits to change the way the counts in the sync file are read.

**Control file**

The control file includes all the necessary parameter to run Bait-ER. The following list describes them:

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Description** | **Notes** |
| Sync\_file | The name of the sync file that we want to analyze. |  |
| Header | This parameter indicates whether the sync file has (value 1) or has not (value 0) a header. |  |
| Columns\_order | The way columns are organized in the sync file. Columns can be organized in time point vs. replicate (value 0) or replicate vs. time point (value 1). | Example. Consider that we have an experimental design with two time points and two replicates. In order 0, we expect the columns to dispose like this: T1R1, T1R2, T2R1 and T2R2. In order 1 we expect the columns to be dispose like this: T1R1, T2R1, T1R2 and T2R2. |
| N\_replicates | The number of replicates |  |
| Time\_points | The vector of time points. Their values should be separated by commas. | Be careful for not introducing any space or other character after or before the comma |
| N\_loci | The number of loci in the sync file. |  |
| Population\_size | The effective population size. | This can be estimated by other methods: e.g. Agnes paper here. |
| Prior\_parameters | The prior parameters alpha and beta of a gamma distribution. These are the parameters of the gamma distribution that models the distribution sigma. | These parameters should be small enough. Simulations conducted by us showed that values smaller or equal than 0.001 are unlike to overdominate the posterior for the generality of E&R experimental designs. |
| Output\_file | The name of the output file. |  |

**Running Bait-ER**

**Once you have your sync and control file, onling with the baiter executable, in the same folder you can now open the termal and run the baiter executable:**

**./baiter**

**Bait-ER will immediatly output some information. Confirm that this information conforms your control file. If not, check whether your control file is correctly filled.**

**A will be printer periodically, so you know how long Bait-ER will take to finish the analyses. However, the output file is updated constantly. If for some reason, your runs are stopped, you can always start by the last site that was printed to the update file.**

**Output file**

The output file has information about the posterior of sigma per locus. The output file has six columns with the following order: chromosome, positions, average sigma, bayes factor for the hypothesis that sigma is different from 0 and the posterior values of alpha a beta.

The output for the first five loci of the example.txt data are is the following:

Rate

Alpha and beta can be used to estimate other quantities of interest regarding the posterior distribution of sigma: quantiles, intervals of confidence, etc. For doing that one can use the gamma function in R. For example, we want a confidence interval at 95% for sigma in position 444, one can simply type in R:

Rate <-

Scale <-

qgamma(0.05,rate,scale) # lower bound

qgamma(0.95,rate,scale) # upper bound

One can conclude that sigma is between [,] with 95% of probability.