

Getting Started

- Step 1. Read SC-Sample paper (see below for details).
- Step 2. Download SC-Sample software.
- Step 3. Compile and run SC-Sample.

Introduction

SC-sample is a computer simulation tool for modeling spatial coalescent with recombination process, which generates ARGs directly consistent with the given sample. It attempts to uniformly select ARG from the sample-consistent ARG space. The algorithm is modified version of SC. The main differences of SC-sample from SC are summarized below:

- 1) There is a new module in the algorithm that judges whether a site is consistent with a local tree (see paper for details);
- 2) Type 1 recombination is found along the sequence as SC. However, instead of a typical exponential distribution, it follows a uniform distribution between current position and the nearest site that is not consistent with the current tree;
- 3) The new lineage's coalescence is also constrained to particular lineages to make sure that the next local tree is nearly consistent with the nearest site that is not consistent with the current tree.

The SC-sample algorithm is described in the following paper:

Ying Wang Ying Zhou, Linfeng Li, Xian Chen, Yuting Liu, Zhi-Ming Ma*, Shuhua Xu*. 2014. ***A New Method for Modeling Coalescent Processes with Recombination***. (Under Review).

Compile SC-Sample software

Requirements:

g++

C++ boost development library (<http://www.boost.org>)

To compile:

make all

Running the SC-sample software

Usage: <samplesize> <region in base pairs> <-f filename> [options]

Options:

-s <random seed>

-d enable debugging messages

-i <iterations>

-h <history> number of previous base pairs to retain

-f <input sample file> sample file name, note that the sample size in given sample file must be same as the value set by <samplesize>

-r <r> (recombination rate per site per 4N generations)

-R <inputfilename> (Tab delimited file where first two columns indicate range of base pair positions scaled to the unit interval and last column is ratio with respect to base line recombination rate.)

-T (Print each local tree in Newick format to standard out)

Example:

```
./sc-sample 20 1000 -r 0.015 -h 1000 -T -f example/sample_10_10.0 -i 100 1>stdout  
2>stderr
```

Input file

Sample file should include SNPs of all chromosomes of the sample, see example/sample_10_10.0 as an example.

Output file

stdout includes all the local tree of the ARGs in the format of newick tree.
stderr includes statistical information of the local trees.