

Getting Started

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

Introduction

SC (Spatial Coalescent simulator) is a computer simulation tool for modeling spatial coalescent with recombination process. The algorithm implemented in SC is an improvement of the Wiuf and Hein algorithm (Wiuf and Hein, 1999). SC is proven to be the only precise ancestral recombination graph (ARG) simulator (Wiuf and Hein's has some redundant and MaCS misses some information) such as the distribution of ARG generated by SC is identical to that generated by a typical back-in-time model adopted by *ms*. SC is developed based on Macs-0.4e, all macs options are not changed. Thanks to Gary Chen to provide support. The main differences of SC from MaCS are listed below:

- 1) Recombination events are divided into two different kinds: type 1 occurs only on the local genealogy at the current position on the sequence; type 2 occurs on other genealogy on the ARG;
- 2) Type 1 recombination is found along the sequence as MaCS, but when the new lineage coalesces to a lineage not on the local genealogy (i.e. a non-ancestral edge), there is an additional judgment to determine whether there is type 2 recombination on that edge. If so, another new lineage appears and the process continues as above until it coalesces to the local genealogy.

These changes make the algorithm substantially more reliable than MaCS and more efficient than the Wiuf and Hein's.

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 software

Requirements:

g++

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

To compile:

make all

Running the SC software

Usage: <samplesize> <region in base pairs> <-r <rate>> <-h <history>> [options]

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

-h <history> number of previous base pairs to retain. It is suggested to use same value as <region in base pairs>, so that SC will

Options:

-s <random seed>
-d enable debugging messages
-i <iterations>
-h <history> number of previous base pairs to retain
-t <mu> (mutation rate per site per 4N generations)
-F <inputfilename> [0|1] (Tab delimited frequency distribution file where first column indicate range of SNP allele frequencies from previous row to current row and last column is desired bin frequency. Second parameter is 1 if SNPs with derived allele freq > 1.0 should have alleles flipped, 0 otherwise.)
-r <r> (recombination rate per site per 4N generations)
-c <f> <lambda> (f = ratio of gene conversion rate to crossover rate. tracklen lambda is mean length of tract in base pairs.)
-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)
-G <alpha> (Assign growth rate alpha across populations where $\alpha = -\log(N_p/N_r)$)
-l <n> <n1> <n2> .. <mig_rate> (Assign all elements of the migration matrix for n populations. Values in matrix set to mig_rate/(n-1).)
-m <i> <j> <m> (Assign i,j-th element of migration matrix to m.)
-ma <m_11>..-n <i> <size> (Pop i has size set to size*N_0)
-g <i> <alpha> (If used must appear after -M option.)
The following options modify parameters at time t.
-eG <t> <alpha> (Assign growth rate for all pops at time t.)
-eg <t> <i> <alpha> (Assign growth rate alpha of pop i at time t.)
-eM <t> <m> (Assign migrate rate m for all elements of migration matrix at time t.)
-em <t> <i> <j> <m_ij> (Assign migration rate for i,j-th element of migration matrix at time t.)
-ema <t> <n> <m_11>..-eN <t> <size> (New pop sizes at time t for all pops where new sizes = size*N_0)
-en <t> <i> <size_i> (New pop size of pop i will be set to (size_i*N_0) at time t.)
-es <t> <i> <p> (Split two populations. At time t, a proportion p of chromosomes from pop i will migrate to a population i+1.)
-ej <t> <i> <j> (Join two populations. At time t all chromosomes migrate from pop i to pop j.)

Example :

./sc 20 167000 -r 6e-4 -h 167000 -T -i 100 1>stdout 2>stderr
The example will generate 100 ARGs with

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 and break points of type2 recombinations(see paper for definition).