Configurations used with program msHOT

María M. Abad-Grau[1] *, Nuria Medina-Medina[1], Rosana Montes-Soldado[1],
José J. Moreno-Ortega[1] and Fuencisla Matesanz[2]

[1] Departamento de Lenguajes y Sistemas Informáticos
Universidad de Granada, Granada 18071, Spain
[2] Instituto de Parasitología y Biomedicina López Neyra
Consejo Superior de Investigaciones Científicas
Granada, Spain

1 Introduction

Recombination fraction within blocks for Africans is 0.000088, which is equivalent to a genetic distance of $0.0088cM$ or a physical distance of $8800bp$. Recombination fraction within blocks for Europeans is 0.000207, which is equivalent to a genetic distance of $0.0207cM$ or a physical distance of $20700bp$. We used Africans.

1.1 Population evolution:

In msHOT evolution is computed backwards from present time.

Africans We want to model a populations which had constant size $N_1 = 10000$ 10000 generations ago and in the next 10000 generations had exponential growth to have a final size $N_2 = N_0 = 1000000$ individuals.

1. Exponential growth: The populations had an exponential growth during $ng = 10000$ generations starting at $N_1 = 10000$ individuals and finishing with $N_2 = N_0 = 1000000$ individuals. This is configured as $-G\ \alpha$, where α is computed by solving equation:

$$N_1 = N_2 exp^{-\alpha ng/(4N_0)},$$

which is $\alpha = 1842.06807$. $-G\ \alpha$ sets growth parameter to α.
2. Constant size: Before that exponential growth, the population was at constant size, thus it was at constant size since 10000 generations ago backwards, which is $eG\ ng/4\ N_0 = 0.0025\ 0.0$

---

* Corresponding author: mabad@ugr.es

Europeans We want to model a populations which had constant size $N_1 = 10000$ 10000 generations ago, then it had an instanteneous size change (bottle neck) and shrinked to $N_2 = 1000$ individuals and in the next 10000 generations had exponential growth to have a final size $N_3 = N_0 = 1000000$ individuals.

1. Exponential growth: The populations had an exponential growth during $ng = 10000$ generations starting at $N_1 = 1000$ individuals and finishing with $N_3 = N_0 = 1000000$ individuals. This is configured as $-G$ α, where α is computed by solving equation:

$$N_1 = N_3 exp^{-\alpha ng/(4N_0)},$$

   which is $\alpha = 2763.1021$. $-G$ α sets growth parameter to α.
2. Constant size: Before that exponential growth, the population was at constant size, thus it was at constant size 10000 generations ago backwards, which is eG $ng/4$ $N_0 = 0.0025$ 0.0
3. Bottle neck: At generation $ng = 10000$ it had an instantaneous size change from 10000 individuals to 1000. This is configured with parameter $-eN$ $ng/(4N_0)$ s, to assign a size of $s \times N_0$ individuals at generation ng (in units of $4N_0$) and a growth rate of 0. Thus, it will be $-eN$ 0.0025 0.01.

1.2   Number of sites

We want to simulate a block of low recombination plus two disease loci, one at the left and the other at the right of the block. We will use 1000 SNPs for each disease locus, in order to increase the chances of having at least 10 SNPs for each disease locus within the frequency interval chosen for the MAF of the disease allele. The number of sites for the block is the number of base pairs, which depends on the recombination factor in the block and thus on the populations. Thus, the total number of sites is 10800 and 22700 for African and European respectively.

1.3   Mutation rate

Mutation rate is $10e - 8$ for both populations. Thus, for all the base pairs is $-t$ $4NxtotalBasePairsx10e - 8$, which is 432 for Africans and 908 for Europeans.

1.4   Basic recombination rate

Recombination rate within the block will be set by using $-r$ $4N_0r \times totalSites$ $totalSites$, which r being the probability of cross-over per generation between the end of the locus being simulated. We will consider r', recombination between consecutive base pairs, to be $10e - 8$ and thus $r = 10e - 8 \times totalBasePairs$. Therefore, it will be $-r$ 432 10800 and $-r$ 908 22700 for Africans and Eurpeans respectively.

1.5   Recombination rates to disease loci

Recombination rate to disease loci is configured as two hotspots in the following way:

$-v$ 2 1000 1001 $recBetween/recPerSiteWithinBlock$ $[nsites-1000]$ $[nsites-999]$ $recBetween2/recWithinBlock$

where $recBetween$ will change to consider different recombination rates to a disease loci. It follows that normal recombination rate $recPerSiteWithinBlock$ was set to $10e-8$, so $recBetween/recPerSiteWithinBlock$ is computed as $recBetween*10e8$. $recBetween2$ is fixed to 0.1, in order to simulate two disase locus not in linkage.