README

MUTATIONAL BURDEN SIMULATOR
with selection, dominance, and demography

Simulation code by David Reich
Documentation by Daniel J. Balick
README v0 created on June 5, 2015

This simulation was used to perform analyses in Do, et al. (2015) and Balick, et al. (2013). Please email dbalick@research.bwh.harvard.edu for questions or comments on this documentation or on the simulator.

Overview

This simulator models population genetics in sexual organisms in a non-equilibrium demography with arbitrary selection strength and dominance coefficients. After initial equilibration, an initial population splits into two sub-populations, one of which experiences a population bottleneck, and the other does not. For representative puurposes we will refer to the bottleneckeded population as "European", and the non-bottlenecked population as "African", as the simulation is particularly useful for modeling the expected genetic differences between Africans and Europeans due to the Out of Africa event. More generally, the simplest model of a square bottleneck is useful for the exploration of parameter space to understand the transient response of alleles under both selection and dominance to a non-equilibrium event such as a bottleneck.

The simulation returns a variety of statistics comparing the mutation burdens of the two populations at a final time.

Simulation details

For computational speed, the simulator only keeps track of allele frequencies in a freely recombining diploid system, rather than containing full genome information. We use an infinite sites model with a mutation rate of 2e-08 per generation per site, with a single selection effect **s** and a single dominance coefficient **h** for all alleles. Allele counts in the current generation are sampled based on the frequencies in the previous generation x_{old}, the selection coefficient s, and the dominance coefficient h. We calculate the

expected frequency x_{current} in the current generation as:

$$x_{current} = \frac{(x_{old}^2(1+s) + x_{old}(1-x_{old})(1+s)h)}{(x_{old}^2(1+s) + 2x_{old}(1-x_{old})(1+s)h + (1-x_{old})^2)}.$$

The simulator has arguments for deleterious mutation rate, ***U***, adding new mutations at a probability of ***U*** per base pair per generation, selection coefficient ***s***, dominance coefficient ***h***, a burn-in of 300,000 generations where sampling occurs every 100 generations in speed-up mode, a transition to sampling every 1 generation at 1000 generations before time ***t***=0.

User specified options

Several arguments are used to run the simulation using the following command.

./SHN_SimpleBN_sim mu s h len burnin1 burnin2 speedup1 speedup2

These values above correspond to the following definitions.

- *mu = argv[1]* is the mutation rate per base pair.
- *s = argv[2]* is the selection coefficient (s < 0 corresponds to deleterious selection).
- *h = argv[3]* is the dominance coefficient (h=0.5 corresponds to additive selection, h=1 corresponds to fully recessive selection).
- *len = argv[4]* is the length of the genome being simulated.
- *burnin1 = argv[5]* is the number of initial burnin generations (for the most ancient burnin).
- *burnin2 = argv[6]* is the number of burnin generations for a more recent burnin.
- *speedup1 = argv[7]* is the speedup factor which is an integer for most ancient burnin.
- *speedup2 = argv[8]* is the speedup factor which is an integer for more recent burnin

The burnin phase can be separated into two epochs allowing for an initial epoch in “speed up” mode (sampling only every *speedup1* generations) and a later burning with more frequent sampling every *speedup2* generation (where speedup2 can be set to 1 if desired).

Simulation output

The simulation computes the mutation burden for each population, given as the sum of all deleterious allele frequencies.

$$\langle x \rangle = \sum_i x_i$$

Comparisons of this quantity between Africans (bottlenecked) and Europeans (bottlenecked) are output in terms of the following statistics.

The burden ratio (R in the output):

$$B_R = \frac{\langle x \rangle_{Afr}}{\langle x \rangle_{Eur}}$$

The burden difference (D in the output):

$$B_D = \langle x \rangle_{Afr} - \langle x \rangle_{Eur}$$

The homozygous burden ratio (R2 in the output):

$$H_R = \frac{\langle x^2 \rangle_{Afr}}{\langle x^2 \rangle_{Eur}}$$

The homozygous burden difference (D2 in the output):

$$H_D = \langle x^2 \rangle_{Afr} - \langle x^2 \rangle_{Eur}$$

In the case of additive selection, the burden ration can be interpreted as the ratio of mutation loads. For completely recessive selection, the homozygous burden ratio can be interpreted as the ratio of mutation loads.

Simulated demographic history

Three basic demographic models are incorporated:

- A simple square bottleneck in Europeans followed by subsequent re-expansion to the original equilibrium population size. Africans rempain in equilibrium throughout.
- The Gravel, et al. (PNAS 2011) model of African and European demographic history as inferred from the 1000 Genomes Project (1KG) data.

- The Tennessen, et al. (Science 2012) model of African and European demogrphic history as inferred from Exome Sequencing Project (ESP) data.

Altering demographic parameters

The code can be easily edited to accommodate custom parameter specifications. In particular, the simple bottleneck simulation can be run over a range of different bottleneck intensities and durations for analysis of the dependence of the mutational burden on the bottleneck intensity and duration.

References

1. Gravel S, et al. (2011) *Demographic history and rare allele sharing among human populations.* Proc. Natl. Acad. Sci. USA 108:11983--11988.
2. Tennessen JA, et al. (2012) *Evolution and functional impact of rare coding variation from deep sequencing of human exomes.* Science 337(6090):64--69.
3. Do R, et al. (2015) *No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans.* Nature Genetics 47:126–131
4. Balick DJ, et al. (2013) *Response to a population bottleneck can be used to infer recessive selection.* http://biorxiv.org/content/early/2014/03/21/003491