Daniel J. Balick's Software and Resources

This page contains links to various software resources associated with active and previous research projects performed with my colleagues. Feel free to email dbalick@hms.harvard.edu or dbalick@gmail.com if you have any questions, comments, or suggestions.

neqPopDynx

This is work in progress, so please use with caution!!. *I will be updating this (with any bug fixes) as needed.*

neqPopDynx (Non-equilibrium Population Dynamix) is a flexible, terminal-based Wright-Fisher simulator scripted in Python 3 for outputting temporal data. Allele frequencies evolve independently in the infinite recombination limit to assess properties of the allele frequency *probability distribution* and are subject to user-specified rates of mutation and back mutation rates (allowing for recurrence), selection and dominance coefficients, initial population size, and a choice of pre-specified demographic changes in the population size with specifiable parameters (e.g., growth rate of exponential expansion, bottleneck start time, duration, and diploid size). The purpose of this simulator is to produce robust temporal output of the non-central moments, central moments, and/or cumulants of the allele frequency probability distribution averaged over L independent sites starting from the same initial frequency $p_0=n/2N$. This allows for comparison to analytic results of the equilibration process towards mutation-selection-drift balance and the dynamics of fully non-equilibrium demographic scenarios (e.g., exponential growth, population bottlenecks).

Files for neqPopDynx_v1.5:

- Python 3 script: neqpopdynx_v1.5.0.py.zip

- Example bash scripts for equilibrium, exponential, bottleneck, and oscillatory demographies: bash_scripts_to_run_neqpopdynx_v1.5.zip

Note: As this is work in progress, if you do happen to use this simulator, please provide any feedback you have via email to dbalick@hms.harvard.edu or dbalick@gmail.com.

simDoSe

Written by Daniel J. Balick

For citations, please reference our American Journal of Human Genetics manuscript.

simDoSe (Simulate Dominance and Selection) is a fast Wright-Fisher simulator for arbitrary diploid selection evolving through realistic human demography.

Features

- Produces a simulated site frequency spectrum (SFS) and summary statistics for user specified demography and diploid selection.
- Models random sampling of a population to output the SFS of a sequenced population sample with user specified sample size.
- Option to create many simulated 'genes' from a single simulation with a larger number of simulated sites.
- Option to create gene sets from an imported list of lengths (i.e., target size/mutation rate)
- Option to simultaneously create 'russian doll' simulated genesets, each formed of genes with descending target size (Lgenes=L/10, Lgenes=L/100, ...)
- Fast and flexible due to the absence of linkage (i.e., infinite recombination limit)
- Arbitrary dominance and selection coefficients, including under- and overdominant diploid selection.
- Properly handles high mutation rates with a command-line option for the recurrent mutation kernel.
- Choose from several literature-based demographies, as well as from equilibrium, linear growth, and exponential growth toy models.
- Can model 'biallelic' genes (e.g., LOF mutations with similar consequence in a single gene)
- Entirely command line-based, so the only needed software is Python 2.7 and the numpy, scipy, and pandas packages.
- Flexible output specification, including full population, population sample, and per-gene site frequency spectra and corresponding summary statistics

Additional details, instructions for running, examples

Please see the simDoSe User Manual for detailed information on the mathematical models and commannd line options.

Downloading simDoSe

simDoSe is available for download on GitHub, where you can find the Python 2.7 script for the latest version, the relevant Anaconda environment shell script, and the User Manual detailing instructions for using simDoSe.

srMLgenes

Written by Daniel M. Jordan
For citations, please reference our American Journal of Human Genetics manuscript.

srMLgenes is a web-based visualization tool to analyze the enrichment of pre-specified or user-uploaded gene sets for strong recessive purifying selection (or for strong additive selection, neutrality, and other diploid selection coefficients of interest).

Features

- Users can view the histogram of the maximum likelihood diploid selection coefficients for a human gene set in the the plane of dominance (h) and selection (s) coefficients.
- Users can view the odds ratio and p-values of the computed enrichment/depletion for various values of selection and dominance relative to the genomic background, viewed in the same space.
- Toggle between (h,s) view and bar plots depicting enrichment for strong recessive and strong additive selection, the primary focus of our comparative analyses.
- Inference was performed using Exome Aggregation Consortium (ExAC) data from the non-Finnish European (NFE) cohort.
- Maximum likelihood values were obtained using simulations performed with simDoSe under a Poisson Random Field (PRF) model.
- Users can explore both ExAC data and simulated gene sets in comparison to simulated genomic backgrounds.
- Gene sets can be restricted to an arbitrary mutational target size (i.e., gene length) range to remove spurious effects from short genes with low confidence inferences

Web link and Download details

The web-based version of srMLgenes is available here and a script can be downloaded directly from GitHub here to run srMLgenes on a local computer.

Mutational Burden Simulator

Written by David Reich

For citations, please reference our PLOS Genetics manuscript and our Nature Genetics manuscript.

Simulating selection and dominance in a non-equilibrium demography

How does a population bottleneck impact the mutation burden under differing dominance and selection coefficients?

Below is the code for **burden_sim**, a simulation of the mutation burden for two populations after a split, as described in Balick et al.

Files for burden_sim

- Readme file: README
- Simulation Code (written in C) burden_sim_code.zip
- Compiled code:
  - Burden simulation for simple (square) bottleneck: burden_sim_simplebn.zip
  - Burden simulation for Gravel demography: burden_sim_gravel.zip
  - Burden simulation for Tennessen demography: burden_sim_tennessen.zip

This page is managed by DJB and does not necessarily reflect the Sunyaev lab as a whole.

From:
https://sunyaevlab.hms.harvard.edu/wiki/!web/ - **Sunyaev Lab**

Permanent link:
https://sunyaevlab.hms.harvard.edu/wiki/!web/dbalick

Last update: **2023/01/09 15:48**