ACTA SCIENTIFIC COMPUTER SCIENCES

Volume 1 Issue 2 December 2019

Short Communication

Island: A Simple Forward Simulation Tool for Population Genetics

Maurice HT Ling*

Colossus Technologies LLP and HOHY Pte Ltd, Singapore

***Corresponding Author:** Maurice HT Ling, Colossus Technologies LLP and HOHY Pte Ltd, Singapore.

Received: November 11, 2019; **Published:** November 28, 2019

Abstract

Changes in population genetic structure can be a result of genetic drift and/or selective pressure, which may result in changes in adaptability of the population. Computer simulations are commonly used to gain insights into the genetic fate of evolving populations. However, most simulation tools in this area require a firm understanding of the mathematical models of genetic drift but low-cost, hands-on tools are the key to make abstract concepts, such as genetic drift, more intuitive. Here, Island is presented as simple forward simulation tool for population genetics based on Mendelian inheritance where a population is generated from a comma-delimited file containing allelic frequencies. Forward simulations start from an initial population and track its evolution over multiple generations. The population is simulated over generations where each generation results in a population file, which can then be examined independently to observe changes in allelic frequencies over generations.

Keywords: Forward Simulation; Population Genetics

Introduction

A common theme in population genetics is to examine the population genetic structure [1]; which is the distribution of genetic variants, also known as genetic diversity, over space and time [2]. Changes in variant frequencies due to randomness is commonly known as genetic drift [3], which has the potential to result in reduction of genetic variation [4]. This differs from selective pressure, which promotes fitter genotypes for survival [5]. The fundamental unit of genetic diversity is an allele [6], which can be defined as a variant form of a gene. Various factors; such as, population size [7] and migration [8]; may result in changes to population genetic variation. Studies have called for the monitoring and maintenance of genetic diversity [9] as it is a necessary component for adaptation [10,11]. However, there has also been cases of increased fitness as a consequence of reduced diversity [12] as a result of genetic purging, which is the reduction or elimination of deleterious allele in the population [13].

As the study of population genetics is primarily observational and monitoring [14], simulation is often used to gain insights into the fate of evolving populations [15,16]. Computer simulation has also been used test predictions on genetic drift in education setting [17] and research setting [18,19]. A number of simulation tools had been developed for population genetics [20] but most require a firm understanding of the mathematical models of genetic drift. This results in substantial learning curve, which is not suitable as an educational tool. Andrews., et al. [21] suggests that 74.6% of biology undergraduates has at least one misconception on genetic drift, underpinning the importance of using simple tools for education. Brewer and Zabinski [22] proposed that low-cost, hands-on tools are the key to make abstract concepts, such as genetic drift, more intuitive to the student. This led to classroom exercises using beans and dice [15], or M&M chocolate candies [23] to teach population genetics. Based on this principle, computational tools are merely a speed-up implementation as understanding population genetics should be underpinned by the fundamental principles of genetic diversity and randomness.

In this communication, a Python-based command-line tool, Island, is presented as a simple forward simulation tool for population genetics based on Mendelian inheritance. Forward simulations start from an initial population and track its evolution under various genetic models, over multiple generations [18]. A population is generated as a file from set of allelic frequencies of various genes provided as a comma-delimited file. The population is then simulated over generations and each generation results in a population file. Each generation can then be examined independently to observe changes in allelic frequencies over generations.

Using island to examine population genetics

Island is implemented as a command-line tool using Python 3 and Python-Fire module (https://github.com/google/python-fire), which aims to simplify the implementation of command-line interface in Python 3. This has been exemplified in previous tools [24,25]. The design of Island is predicated on simplicity of use. There are three steps to using Island to examine population genetics.

Firstly, a population is generated from a comma-delimited file containing allelic frequencies using gpop command. For

example, the command-line statement, `python island.py gpop --populationfile=test_pop --ploidy=2 --generation_count=0 --population_size=10 --parameterfile=island_parameter.csv`; takes the allelic frequency file, island_parameter.csv, to generate the population as a file, test_pop, of 10 organisms (population_size=10) where each organism is a diploid (ploidy=2) and the current generation count is zero to denote ancestral generation. The organisms are identified by an incremental number within the generation. The generation count is not used in the simulation but improves documentation of the simulation.

The format of the parameter file of allelic frequencies is as follows: (a) A comma-delimited header row of gene name and allelic names but since the allelic names of each gene are different, the allelic names tend to be a running symbol of "AF1, AF2, ..." to denote allelic frequency 1 and allelic frequency 2, respectively. (b) The allelic frequencies of each gene are represented as a comma-delimited data row of gene name, followed by allelic frequencies. For example, `ACTB,0.5,0.25,0.25` represents 50%, 25%, and 25% of ACTB allele 1, 2, and 3, respectively. The names of alleles are not used in Island; instead, each allele of a gene is denoted by running incremental order of appearance.

Secondly, a population is simulated using simulate command. For example, `python island.py simulate --populationfile=test_pop --simulation_type=simple --population_size=10 --generations=10`; takes the population file test_pop to run a simple simulation (simulation_type=simple) for 10 generations (generations=10) where the population size at the end of each generation is fixed at 10 organisms (population_size=10). Currently, there is only simple simulation, which is one that (a) assumes diploid organisms, (b) one random crossover per chromosome pair, (c) crossover is generated prior to mating to simulate random haploid, (d) no mutation, (e) random mating without possibility of self-mating, (f) mating only within generation, and (g) population size may be changed. The change of population size is to cater for population bottleneck [26,27] and the rate of its subsequent recovery [28]. Each generation will generate a population file with incremental number; for example, in this case, test_pop.1 and test_pop.10 files will represent the population of the first and 10th generation of simulation respectively; which can then be used initiate further simulations if necessary.

Finally, each generation, represented by its population file, can be analysed independently and collated to examine changes in allelic frequencies over generations. Currently, only tabulation of allelic counts is available. The command, `python island.py tabulateCount --ploidy=2 --populationfile=test_pop;` analyses the population in test_pop file where each organism is a diploid (ploidy=2) to generate a table of expected and observed allelic counts for further statistical analyses.

Based on these steps, Island can be used to examine several interesting hypotheses in population genetics; such as,

- Act as control (genetic drift) simulations against simulations with selective pressure.
- Effects of initial population genetic diversity in subsequent generations.
- Effects of population size in subsequent genetic diversity.
- Effects of gene flow [29] to populations of different diversities by combining two population files using combinepop command. For example, `python island.py combinepop --populationfile1=test_pop.2 --populationfile2=test_pop.3 --outputfile=test_pop.comb`, combines the organisms in test_pop.2 (populationfile1=test_pop.2) and test_pop.3 (populationfile2=test_pop.3) into test_pop.comb (outputfile=test_pop.comb) where the organisms will be renumbered.

Conclusion and future work

In summary, this communication presents the structure of Island and its potential use in forward simulation of population genetics [18] from a simple description file of allelic frequencies. This is entirely based on Mendelian inheritance. Future work can expand on various simulation scenarios previously reviewed [20] and cater to non-Mendelian genetics, as well as interfacing with other simulation tools for selective pressure; such as, Avida [30] and DOSE [8]; for mixed genetic drift / selective pressure simulations. Moreover, additional analytical operations such as coefficient of inbreeding [31] and fixation index [32], can be added.

Availability

Island is part of the Bactome project (https://github.com/mauriceling/bactome) and is licensed under GNU General Public Licence version 3 for academic and non-commercial purposes only.

Conflict of Interest

The author declares no conflict of interest.

Bibliography

1. JH Sul., *et al*. "Population structure in genetic studies: Confounding factors and mixed models". *PLoS Genetics* 14.12 (2018): e1007309.

2. R F Hoekstra. "Population genetics of filamentous fungi". *Antonie Van Leeuwenhoek* 65.3 (1994): 199–204.

3. J Masel. "Genetic drift". *Current Biology* 21.20 (2011): R837-838.

4. B Star and HG Spencer. "Effects of genetic drift and gene flow on the selective maintenance of genetic variation". *Genetics* 194.1 (2013): 235-244.

5. M Ueda., *et al*. "Stronger selection can slow down evolution driven by recombination on a smooth fitness landscape". *PLoS ONE* 12.8 (2017): e0183120.

6. W Arthofer., *et al*. "Identifying the minimum number of microsatellite loci needed to assess population genetic structure: A case study in fly culturing". *Fly (Austin)* 12.1 (2018): 13-22.

7. RC Lacy. "Importance of genetic variation to the viability of mammalian populations". *Journal of Mammalogy* 78.2 (1997): 320-335.

8. CF Castillo and MH Ling. "Digital Organism Simulation Environment (DOSE): a library for ecologically-based in silico experimental evolution". *Advances in Computer Science: an International Journal* 3.1 (2014): 44-50.

9. M Govindaraj., *et al*. "Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives". *Genetics Research International* (2015): 1-14.

10. YT Lai., *et al*. "Standing genetic variation as the predominant source for adaptation of a songbird". *Proceedings of the National Academy of Sciences of the United States of America* 116.6 (2019): 2152.

11. K Schmidt., *et al*. "The importance of genetic variability and population differentiation in the Eurasian lynx Lynx lynx for conservation, in the context of habitat and climate change". *Mammal Review* 41.2 (2011): 112-124.

12. C Dale and NA Moran. "Molecular interactions between bacterial symbionts and their hosts". *Cell* 126.3 (2006): 453-465.

13. A García-Dorado. "On the consequences of ignoring purging on genetic recommendations for minimum viable population rules". *Heredity (Edinb)* 115.3 (2015): 185-187.

14. G Leroy., *et al*. "Next-generation metrics for monitoring genetic erosion within populations of conservation concern". *Evolutionary Applications* 11.7 (2018): 1066-1083.

15. CAM Russo and CM Voloch. "Beads and dice in a genetic drift exercise". *Evolution: Education and Outreach* 5.3 (2012): 494-500.

16. NR Nené., *et al*. "Evaluating genetic drift in time-series evolutionary analysis". *Journal of Theoretical Biology* 437 (2018): 51-57.

17. RM Price., *et al*. "Observing populations and testing predictions about genetic drift in a computer simulation improves college students' conceptual understanding". *Evolution: Education and Outreach* 9.1 (2016): 8.

18. X Yuan., *et al*. "An overview of population genetic data simulation". *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 19.1 (2012): 42-54.

19. P A Saunders., *et al*. "Sex chromosome turnovers and genetic drift: a simulation study". *Journal of Evolutionary Biology* 31.9 (2018): 1413-1419.

20. S Hoban., *et al*. "Computer simulations: tools for population and evolutionary genetics". *Nature Reviews Genetics* 13.2 (2012): 110-122.

21. TM Andrews., *et al*. "Biology undergraduates' misconceptions about genetic drift". *CBE Life Sciences Education* 11.3 (2012): 248-259.

22. CA Brewer and C Zabinski. "Simulating genetic change in a large lecture hall: the ultimate bean counting experience". *The American Biology Teacher* 61.4 (1999): 298-302.

23. NL Staub., *et al*. "Teaching evolutionary mechanisms: genetic drift and M&M's®". *Bio Science* 52.4 (2002): 373-377.

24. MH Ling. "SEcured REcorder BOx (SEREBO) based on blockchain technology for immutable data management and notarization". *MOJ Proteomics and Bioinformatics* 7.6 (2018): 169-174.

25. MH Ling. "RANDOMSEQ: Python command–line random sequence generator". *MOJ Proteomics and Bioinformatics*7.4 (2018): 206-208.

26. W Wang., *et al*. "Low genetic diversity in a critically endangered primate: shallow evolutionary history or recent population bottleneck?". *BMC Evolutionary Biology* 19.1 (2019): 134, 26.

27. M Ørsted., *et al*. "Genomic variation predicts adaptive evolutionary responses better than population bottleneck history". *PLoS Genetics* 15.6 (2019): e1008205.

28. KM Cammen., *et al*. "Genomic signatures of population bottleneck and recovery in Northwest Atlantic pinnipeds". *Ecology and Evolution* 8.13 (2018): 6599-6614.

29. LS Miles., *et al*. "Gene flow and genetic drift in urban environments". *Molecular Ecology* 28.18 (2019): 4138-4151.

30. C Ofria and CO Wilke. "Avida: a software platform for research in computational evolutionary biology". *Artificial Life* 10.2 (2004): 191-229.

31. S Wright. "Coefficients of inbreeding and relationship". *The American Naturalist* 56.645 (1922): 330-338.

32. KE Holsinger and BS Weir. "Genetics in geographically structured populations: defining, estimating and interpreting F(ST)". *Nature Reviews Genetics* 10.9 (2009): 639-650.

Volume 1 Issue 2 December 2019