

# Computer Note

## EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations

F. Balloux

If there are many programs for population genetics data analysis, less effort has been devoted to simulation software in this field. However, there are many situations where simulated population genetics datasets would be very useful. Simulations allow exploring situations too complex to be solved analytically. Indeed, outside very simple population genetics models, it is often extremely difficult to obtain the expectations for quantities such as gene correlation or effective size. Simulations also permit us to validate inferences drawn from empirical studies. They can be run with different estimated parameters and compared to the original dataset. It is also possible to build any null hypotheses against which inferences drawn from real data can be tested. Another nice feature of simulated datasets is the possibility of testing new software devoted to population genetics analysis under nontrivial conditions. In addition to their value in research, simulated data can be very useful for educational purposes. Various datasets can be generated and provided to students to train their skills in population genetics data analysis. All these applications prompted me to develop the software EASYPOP, an individual-based model that simulates neutral loci datasets under a very broad range of conditions.

### Program Description

EASYPOP can simulate haploid, diploid, or haplodiploid organisms. For haploid organisms, the proportion of clonal and sexual reproduction (recombination) can be set, whereas for diploid organisms with

one sex, the proportion of selfing can also be selected. For diploid organisms with two sexes, there is the choice between hermaphrodites or sexuals. For hermaphrodites, the proportion of selfing can be set, whereas for sexuals complex breeding structures can be simulated (e.g., monogamy with a given proportion of extra-pair matings). Numbers of individuals (males and females) can be set independently for each population and dispersal is sex specific. Various migration models are implemented, such as the two-dimensional stepping stone or hierarchical island model. In addition, there is an isolation by distance option that works with the coordinates of the populations in any number of dimensions. The migration model and migration rates can be modified after a certain number of generations. This later option allows us to explore the genetic consequences of a modification of the migration pattern or to impose a bottleneck on a population. There are also several mutation models implemented, which are oriented toward the simulation of microsatellite loci. Mutation rates, mutation models, and the maximal number of allelic states can be set individually for each locus. Genotypes are real multilocus; that is, genotypes are not independent replicates over loci.

EASYPOP can handle very large simulations on standard personal computers. Most parameters are limited only by the memory of the computer. However, the number of possible allelic states and the number of populations are limited to 999 and 10,000, respectively. It is possible to run up to 99 independent replicates. The program works with probabilities (e.g., migration, mutation), and thus allows fitting to analytical variance expectations.

EASYPOP has been debugged as carefully as possible by testing simulation results against analytical expectations in all cases where they could be computed.

Some of these tests can be found in Balloux et al. (2000).

### Input and Output

The input consists only of the parameters of the simulation chosen by the user. The outputs of the program are the multilocus genotypes after a given number of generations. Output files fitting FSTAT (Goudet 1995), GENEPOP (Raymond and Rousset 1995) or ARLEQUIN (Schneider et al. 1997) are generated. This allows for direct analysis of the results using the mentioned population genetics packages. An additional output file giving the number of alleles, Nei and Chesser's (1983) unbiased heterozygosities, and  $F$  statistics for each 10th generation is provided. This file can be used to follow these quantities over time.

### How to get the Program

EASYPOP is available in Mac PPC and Windows 95/98/NT4 versions. The program can be downloaded free of charge with its user manual at the following website: <http://www.unil.ch/izea/software/easypop.html>. After each substantial development of EASYPOP, people who register the program will be contacted about updates. The C source code can be obtained by e-mailing: [francois.balloux@esh.unibe.ch](mailto:francois.balloux@esh.unibe.ch).

From the Institute of Zoology, University of Bern, CH-3032, Hinterkappelen-Bern, Switzerland. It is a pleasure to thank Jean Lehman who introduced me to programming. I am also deeply indebted to Max Reuter who showed me how to program in C, and helped me numerous times to find out some vicious bugs. I also thank Jérôme Goudet, Nicolas Perrin, Eric Petit, Thierry De Meeüs, Pierre Fontanillas, Laurent Lehman, Patrick Basset, and Renaud Kaeuffer for the time they invested to test this program and for comments and suggestions. Address correspondence to Francois Balloux at the address above or e-mail: [francois.balloux@esh.unibe.ch](mailto:francois.balloux@esh.unibe.ch).

© 2001 The American Genetic Association

## References

- Balloux F, Brüner H, Lugon-Moulin N, Hausser J, and Goudet J, 2000. Microsatellites can be misleading: an empirical and simulation study. *Evolution* 54:1414–1422.
- Goudet J, 1995. FSTAT: A computer program to calculate F-statistics. *J Hered* 86:485–486.
- Nei M and Chesser RK, 1983. Estimation of fixation indices and gene diversities. *Ann Hum Genet* 47:253–259.
- Raymond M and Rousset F, 1995. GENEPOP: population genetics software for exact tests and ecumenicism. *J Hered* 86:248–249.
- Schneider S, Kueffer JM, Roessli D, and Excoffier L, 1997. ARLEQUIN (version 1.1): a software for population genetic data analysis. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Received June 14, 2000

Accepted October 31, 2000

Corresponding Editor: Sudhir Kumar