

Genetics and population analysis

quantiNemo: an individual-based program to simulate

data, citation and similar papers at core.ac.uk

brought to

provided by RERC

metapopulation

Samuel Neuenschwander^{1,*}, Frédéric Hospital^{1,2}, Frédéric Guillaume³ and Jérôme Goudet¹¹Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland, ²INRA, UMR Génétique et Diversité Animales, 78352 Jouy-en-Josas Cedex, France and ³Department of Zoology, University of British Columbia, Vancouver, B.C., V6T 1Z4, Canada

Received on March 15, 2008; revised on April 25, 2008; accepted on April 30, 2008

Advance Access publication May 1, 2008

Associate Editor: Martin Bishop

ABSTRACT

Summary: quantiNemo is an individual-based, genetically explicit stochastic simulation program. It was developed to investigate the effects of selection, mutation, recombination and drift on quantitative traits with varying architectures in structured populations connected by migration and located in a heterogeneous habitat. quantiNemo is highly flexible at various levels: population, selection, trait(s) architecture, genetic map for QTL and/or markers, environment, demography, mating system, etc. quantiNemo is coded in C++ using an object-oriented approach and runs on any computer platform.

Availability: Executables for several platforms, user's manual, and source code are freely available under the GNU General Public License at <http://www2.unil.ch/popgen/softwares/quantinemo>

Contact: samuel.neuenschwander@unil.ch

1 INTRODUCTION

Selection plays a major role in the evolution of quantitative traits and thus on the evolution of organisms. Understanding the effects of selection is essential for population and quantitative geneticists. However, the effects of selection on quantitative traits are complex, interfere with other evolutionary forces (drift, migration, mutation and recombination) and are not well understood. Simulating the evolution of quantitative traits with explicit architecture in an individual-based, forward time model could help to address such questions, as shown by recent launch of simulation programs like EasyPop (Balloux, 2001), FPG (Hey, 2004), simuPOP (Peng and Kimmel, 2005) or NEMO (Guillaume and Rougemont, 2006). However, none of these programs allow investigating the evolution of complex quantitative traits under selective pressures in a realistic environment. Here we present quantiNemo, an individual-based, genetically explicit stochastic simulation program, which allows the simulation of genetic and phenotypic data for complex quantitative traits in a realistic and dynamic metapopulation with selective pressures that can vary in space and time. The loci underlying the traits and the markers can

be placed on a genetic map, allowing the simulation of realistic genetic maps. The implementation of selection is highly flexible. quantiNemo is built in the evolutionary and population genetics-programming framework NEMO (Guillaume and Rougemont, 2006).

2 FEATURES

quantiNemo consists of several simulation components, which may be easily extended. Here we describe the major components available in quantiNemo. A full description of quantiNemo functionalities and parameters can be found in the user's manual.

Quantitative traits: quantiNemo allows the simulation of multiple quantitative traits each having its own specifications. Each trait is defined by one to many loci each with up to 256 alleles. The allelic effects at each locus can be drawn from a normal distribution or can be set explicitly. Mutations are implemented with several models (Random, Incremental) or could be explicitly defined. The trait determinism can be purely additive, or include dominance and/or epistatic interactions among loci. Environmental effects can also be set in different ways. Currently, quantiNemo does not allow specifying correlated traits or trade-offs among traits, but this could be implemented in a future release.

Neutral markers: quantiNemo allows the simulation of neutral markers, such as microsatellites or single nucleotide polymorphisms (SNP) with different mutation models (K Allele, Stepwise).

Different types of neutral markers and/or quantitative trait loci can be combined within the same simulation.

Genetic map: Markers and QTLs can be located on a genetic map, which may consist of several chromosomes. This allows an explicit positioning of all types of loci on the map.

Metapopulation: quantiNemo allows simulating realistic population dynamics. The carrying capacities of populations can vary in space and time, and population sizes can be fixed or determined by selective or other forces. The user can use several preset migration models (island, 1-D stepping-stone, 2-D stepping-stone), or specify the full migration matrix. The migration pattern can change over

* To whom correspondence should be addressed.

time, allowing investigating scenarios of population fragmentation or invasion of new habitats.

Lifecycle: Each individual undergoes a single life cycle (non-overlapping generations). The life cycle is fixed [in contrast to NEMO (Guillaume and Rougemont, 2006)] and starts with breeding and reproduction. Several mating systems are available: random mating or selfing for hermaphrodites, promiscuity, monogamy or polygyny for species with separate sexes. Selection acts on the reproductive fitness of individuals. After reproduction, juveniles may disperse to other populations, and then population size is possibly regulated. Environmental stochasticity can also be introduced, where populations may go extinct due to an external factor, independent of population size or genetic constitution of the population.

Selection: Several modes of selection are available which may differ among quantitative traits. In stabilizing selection modes, a specific optimum and selection intensity may be defined for each population and quantitative trait (Bürger, 2000). In directional selection modes, the strength and direction of selection may vary for each trait and population. Furthermore, the selective pressures can change over time. Last, selection can be soft or hard.

Input: quantiNemo is launched using a settings file. The settings file is a text file with a flexible and easy to understand structure, can be edited with any text editor, and comments may be added for better readability.

Output: quantiNemo provides summary statistics for the different simulation components, including genetic variance estimates, quantitative trait analysis (e.g. Q_{ST}), and F -statistics for all types of loci (neutral and QTL). Most of the summary statistics are available for juveniles and adults. The summary statistics can be computed for any generation during the simulation.

quantiNemo can also produce files with the raw genetic data. The genotypes at all loci can be dumped to file in the FSTAT format (Goudet, 1995). Phenotypes, genotypes, as well as additive, dominance and epistatic effect values can be dumped to a file and then analyzed with any population or quantitative genetic software, for instance to study linkage disequilibrium, or scan for the presence and position of QTLs. We are currently developing an R package to carry out the most common analyses from a simulation.

Technical: quantiNemo is a console program, and is coded in standard C++ using an object-oriented approach. This allows compiling quantiNemo on any computer platform, which supports standard C++ compilation. There is no limit on the number of populations, individuals, genes, etc. that quantiNemo can handle, apart from the available hardware capacities (CPU and memory). quantiNemo was optimized for high computation efficiency in particular for large simulations on clusters. For instance, the simulations shown in Figure 1 including a total of 1.5×10^9 items (loci \times individuals \times pop \times generations) lasted <9 min on a 3.2-GHz processor running Linux.

The website <http://www2.unil.ch/popgen/software/quantinemo> includes executables for Windows, Linux and Mac, the source code, a detailed user's manual and also a syntax highlighting definition to edit the settings file with the shareware TextPad (<http://www.textpad.com>). All are freely available under the terms of the GNU General Public License.

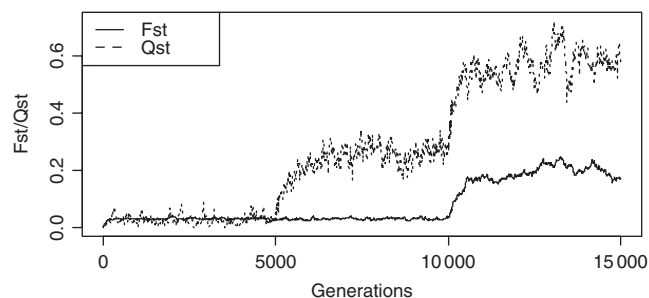


Fig. 1. Example of the evolution of genetic differentiation between five populations (each made of 1000 individuals) over 15 000 generations. The figure shows how differentiation at a quantitative trait (Q_{ST}) determined by 10 QTLs and at 10 unlinked neutral markers (F_{ST}) change under the influence of selection and migration. During the first 5000 generations the quantitative trait evolved neutrally. This period is followed by selection for different local optima in the different populations ($V_{Zopt} = 2.5$, $w = 10$). At generation 10 000 gene flow between the populations was reduced by a factor 10 (from 5 to 0.5 emigrants per generation and population).

3 DISCUSSION

The availability of quantitative traits and their QTLs, neutral markers, genetic map and structured populations in a single program makes quantiNemo ideally suited for a wide range of applications, from population genetics, evolutionary quantitative genetics, to association studies and QTL mapping. For instance Figure 1 shows how selection and gene flow influence genetic differentiation at neutral markers and for a phenotypic trait submitted to different selective pressures.

Finally, quantiNemo has a high educational value. Because quantiNemo is designed as an end-user simulation program with a flexible and clear settings file, everybody and students in particular can perform simulations (Fig. 1) and gain many insights by visualizing the joint effects of selection, linkage, migration, etc.

Last, the object-oriented design of quantiNemo facilitates the implementation of additional functionalities by more expert users.

ACKNOWLEDGEMENTS

Funding: This work has been financially supported by a Swiss NSF grant no. 3100A0-108194/1 to J.G. and a Herbertte Foundation grant (University of Lausanne) to J.G. and F.H.

Conflict of Interest: none declared.

REFERENCES

- Balloux,F. (2001) EASYPOP (Version 1.7): a computer program for population genetics simulations. *J. Her.*, **92**, 301–302.
- Bürger,R. (2000) *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Cichester, UK.
- Goudet,J. (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *J. Her.*, **86**, 485–486.
- Guillaume,F. and Rougemont,J. (2006) Nemo: an evolutionary and population genetics programming framework. *Bioinformatics*, **22**, 2556–2557.
- Hey,J. (2004) FPG: a computer program for forward population genetic simulation. Available at <http://lifesci.rutgers.edu/~hey/lab/HeylabSoftware.htm#FPG>.
- Peng,B. and Kimmel,M. (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21**, 3686–3687.