BIOINFORMATICS APPLICATIONS NOTE

Genetics and population analysis

SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination

Nicolas Ray1,†, Mathias Currat2,†,⋆, Matthieu Foll3,4 and Laurent Excoffier3,4

1EnviroSPACE laboratory, Climate Change and Climate Impacts, Institute for Environmental Sciences, University of Geneva, Battelle - Building D, 7 route de Drize, 1227 Carouge, 2Laboratory of Anthropology, Genetics and Peopling History, Department of Anthropology, University of Geneva, 1227 Geneva, 3Computational and Molecular Population Genetics Lab (CMPG), Institute of Ecology and Evolution, University of Berne, Baltzerstrasse 6, 3012 Berne and 4Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Associate Editor: Jeffrey Barrett

ABSTRACT
Summary: SPLATCHE is a program to simulate the demography of populations and the resulting molecular diversity for a wide range of evolutionary scenarios. The spatially explicit simulation framework can account for environmental heterogeneity and fluctuations, and it can manage multiple population sources. A coalescent-based approach is used to generate genetic markers mostly used in population genetics studies (DNA sequences, SNPs, STRs or RFLPs). Various combinations of independent, fully or partially linked genetic markers can be produced under a recombination model based on the ancestral recombination graph. Competition between two populations (or species) can also be simulated with user-defined levels of admixture between the two populations. SPLATCHE2 may be used to generate the expected genetic diversity under complex demographic scenarios and can thus serve to test null hypotheses. For model parameter estimation, SPLATCHE2 can easily be integrated into an Approximate Bayesian Computation (ABC) framework.

Availability and implementation: SPLATCHE2 is a C++ program compiled for Windows and Linux platforms. It is freely available at www.splatche.com, together with its related documentation and example data.

Contact: mathias.currat@unige.ch

Received on August 19, 2010; revised on September 23, 2010; accepted on October 11, 2010

1 INTRODUCTION

SPLATCHE2 (for SPatiaL And Temporal Coalescences in Heterogeneous Environments, version 2.0) is derived from the program SPLATCHE released in 2004 (Currat et al., 2004). This coalescent-based program has been developed to model complex and spatially explicit demographic scenarios, such as range expansions, in a heterogeneous and dynamic environment, and to simulate the resulting molecular diversity of sampled individuals at arbitrary locations. It thus makes it possible to realistically model the environment in order to study the impact of ecological factors on the genetic structure of populations. This approach has already proved invaluable to study the patterns of genetic variation in spatially explicit contexts (Foll and Gaggiotti, 2006; Ray et al., 2005) or the genetic consequences of range expansion (Currat et al., 2006; Hamilton et al., 2005; Klopfstein et al., 2005; Ray et al., 2003; Wegmann et al., 2006).

Many improvements and new functionalities have been added here as compared to the original version of SPLATCHE, allowing the simulation of a much wider range of evolutionary scenarios. Some of these new features have already been used in published studies, such as the simulation of admixture and competition between populations (Currat and Excoffier, 2004, 2005; Currat et al., 2008; Francois et al., 2010) or the possibility of defining multiple source populations (Ray et al., 2005). Those extensions have now been incorporated into SPLATCHE2, together with the ability (i) to simulate linked or partially linked loci, (ii) to have multiple coalescent events per generation and (iii) to generate single-nucleotide polymorphism (SNP). Moreover, SPLATCHE2 is also available as a non-graphical console program that can be used on computer clusters. Here, we briefly describe the main features of SPLATCHE2, noting that a more detailed description can be found in a user manual.

2 METHODS AND IMPLEMENTATION

2.1 Demographic simulation

SPLATCHE2 begins by simulating the demography of one (or two overlapping and interacting) subdivided population(s) made up of haploid individuals. The population demes are arranged on a 2D lattice and each population can expand from one or several sources. Local deme density is logistically regulated, and migration can occur between a deme and its four nearest neighbors. Directional constraints for migration can be defined through a friction parameter. Demes carrying capacity and friction can be set to arbitrary values in every deme and changed at any generation to simulate environmental heterogeneity and its fluctuation over time. Competition between two populations (or species) can also be simulated, potentially leading to the extinction of one of them. An admixture parameter λ may be used to simulate gene flow between the two populations (or species). The intensity of inter-population gene flow can vary between 0 (no admixture) and 1 (full interbreeding).

2.2 Coalescent simulation

The second phase of the simulations uses the demographic history simulated in the previous forward step (population sizes, admixture and migration events) to generate the genetic diversity for a set of samples drawn from the
population. A backward, discrete coalescent model (Hudson, 1990; Kingman, 1982) including recombination based on the ancestral recombination graph (as in the program SIMCOAL2, Laval and Excoffier, 2004) is implemented. Under the coalescent approach, only the genealogy of sampled lineages needs to be reconstructed, which results in an enormous gain in term of computation time.

The coalescent tree obtained after such a simulation is then used to generate molecular diversity for the set of samples. Various commonly used genetic markers can be simulated, such as full DNA sequences, SNPs, short tandem repeats (STRs) and restriction fragment length polymorphism (RFLP). Various combinations of independent, fully or partially linked loci may be simulated. Note that despite simulating haploid genes, diploid individuals can be simulated by combining pairs of alleles or chromosomes under the assumption of Hardy–Weinberg equilibrium.

Other coalescent simulation programs can generate genetic data under an island model (Hudson, 2002; Laval and Excoffier, 2004). Moreover, other forward simulation programs can also generate data under complex demographic models (Neuenschwander et al., 2008). To our knowledge, SPLATCHE2 is currently the only coalescent simulation program that can integrate a dynamic and spatially explicit environment.

2.3 Outputs and integration into an approximate Bayesian computation framework

The main outputs of SPLATCHE2 are a set of ARLEQUIN files (Excoffier et al., 2005; Excoffier and Lischer, 2010) including the resulting genetic diversity of the simulated samples. These files can thus be analyzed by ARLEQUIN to compute a wide array of summary statistics on the data. Moreover, many aspects of the coalescent process may be directly visualized, either as a tree (NEXUS format) or as the geographic locations of coalescent events. When simulating two interbreeding populations, one can output the resulting final levels of genetic introgression between populations. When simulating two interbreeding populations, one can output either as a tree (NEXUS format) or as the geographic locations of coalescent events. When simulating two interbreeding populations, one can output the resulting final levels of genetic introgression between populations.

Moreover, many aspects of the coalescent process may be directly visualized, either as a tree (NEXUS format) or as the geographic locations of coalescent events. When simulating two interbreeding populations, one can output the resulting final levels of genetic introgression between populations.