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

Nicolas Ray[1,†], Mathias Currat[2,†,*], Matthieu Foll[3,4] and Laurent Excoffier[3,4]

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

Associate Editor: Jeffrey Barrett

ABSTRACT

Summary: SPLATCHE2 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

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

---

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

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 (SNPs). 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. Deme 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 (Currat *et al.*, 2008). In the Windows graphical interface, various aspects of the demography and genetics can also be visualized (e.g. bitmaps of deme densities, migration intensities, movement of genes during the backward simulation). The command-line version of SPLATCHE2 has been developed to facilitate the parallelization of the simulations, such as to allow one to estimate population parameters [e.g. Approximate Bayesian Computation (ABC); Beaumont *et al.*, 2002]. Indeed, SPLATCHE2 can be easily integrated into the ABCtoolbox framework (Wegmann *et al.*, 2010) allowing one to get posterior distributions of various demographic and genetic parameters.

3 CONCLUSION

SPLATCHE2 can be used to simulate complex and realistic demographic models and to generate the associated molecular diversity of sampled individuals. Alternative population or environmental histories can be modelled and compared through their impacts on resulting genetic diversity. Due to its explicit handling of spatial information and of environmental and temporal heterogeneities, SPLATCHE2 is particularly well suited for studying spatially distributed population samples over relatively short evolutionary time scales (i.e. a few thousand generations).

REFERENCES

Beaumont,M.A. *et al.* (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Currat,M. and Excoffier,L. (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol.*, **2**, e421.

Currat,M. and Excoffier,L. (2005) The effect of the Neolithic expansion on European molecular diversity. *Proc. Roy. Soc. Lond B. Biol. Sci.*, **272**, 679–688.

Currat,M. *et al.* (2006) Comment on "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens" and "Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans". *Science*, **313**, 172; author reply 172.

Currat,M. *et al.* (2004) SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Mol. Ecol. Notes*, **4**, 139–142.

Currat,M. *et al.* (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–1920.

Excoffier,L. *et al.* (2005) ARLEQUIN (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinformatics Online*, **1**, 47–50.

Excoffier,L. and Lischer,H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.*, **10**, 564–567.

Foll,M. and Gaggiotti,O. (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–891.

Francois,O. *et al.* (2010) Principal component analysis under population genetic models of range expansion and admixture. *Mol. Biol. Evol.*, **27**, 1257–1268

Hamilton,G. *et al.* (2005) Bayesian estimation of recent migration rates after a spatial expansion. *Genetics*, **170**, 409–417

Hudson,R.R. (1990) *Gene Genealogies and the Coalescent Process*. Oxford University Press, Oxford.

Hudson,R.R. (2002) Generating samples under a Wright-Fisher neutral model. *Bioinformatics,* **18**, 337–338

Kingman,J.F.C. (1982) The coalescent. *Stoch. Proc. Appl.*, **13**, 235–248.

Klopfstein,S. *et al.* (2005) The fate of mutations surfing on the wave of a range expansion. *Mol. Biol. Evol.*, **23**, 482–490

Laval,G. and Excoffier,L. (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, **20**, 1485–2487.

Neuenschwander,S. *et al.* (2008) quantiNEMO: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics* **24**, 1552–1553.

Ray,N. *et al.* (2005) Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Res.*, **15**, 1161–1167.

Ray,N. *et al.* (2003) Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.*, **20**, 76–86.

Wegmann,D. *et al.* (2006) Molecular diversity after a range expansion in heterogeneous environments. *Genetics*, **174**, 2009–2020.

Wegmann,D. *et al.* (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 116.