



Article scientifique

Article

2019

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

SPLATCHE3: simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal

Currat, Mathias; Arenas, Miguel; Quilodran, Claudio; Excoffier, Laurent; Ray, Nicolas

How to cite

CURRAT, Mathias et al. SPLATCHE3: simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal. In: *Bioinformatics*, 2019, vol. 35, n° 21, p. 4480–4483. doi: [10.1093/bioinformatics/btz311](https://doi.org/10.1093/bioinformatics/btz311)

This publication URL: <https://archive-ouverte.unige.ch/unige:117770>

Publication DOI: [10.1093/bioinformatics/btz311](https://doi.org/10.1093/bioinformatics/btz311)

Application notes

SPLATCHE3: simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal

Mathias Currat^{1,2*}, Miguel Arenas^{3,4}, Claudio S. Quilodran¹, Laurent Excoffier^{5,6°}, Nicolas Ray^{7,8°}

¹ Laboratory of anthropology, genetics and peopling history, Department of Genetics and Evolution – Anthropology Unit, University of Geneva, 1205 Geneva, Switzerland, ² Institute of Genetics and Genomics in Geneva (IGE3), University of Geneva, 1211 Geneva, Switzerland, ³ Department of Biochemistry, Genetics and Immunology, University of Vigo, 36310 Vigo, Spain, ⁴ Biomedical Research Center (CINBIO), University of Vigo, 36310 Vigo, Spain, ⁵ Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland, ⁶ Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland, ⁷ Institute of Global Health, GeoHealth group, University of Geneva, 1205 Geneva, Switzerland, ⁸ Institute for Environmental Sciences, University of Geneva, 1205 Geneva, Switzerland.

*To whom correspondence should be addressed.

°These authors contributed equally to this work.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: SPLATCHE3 simulates genetic data under a variety of spatially explicit evolutionary scenarios, extending previous versions of the framework. The new capabilities include long-distance migration, spatially and temporally heterogeneous short-scale migrations, alternative hybridization models, simulation of serial samples of genetic data and a large variety of DNA mutation models. These implementations have been applied independently to various studies, but grouped together in the current version.

Availability and Implementation: SPLATCHE3 is written in C++ and is freely available for non-commercial use from the website <http://www.splatche.com/splatche3>. It includes console versions for Linux, MacOs and Windows and a user-friendly GUI for Windows, as well as detailed documentation and ready-to-use examples.

Contact: mathias.currat@unige.ch

Supplementary information: Details about the implemented evolutionary models and parameters are provided in the software documentation available at <http://www.splatche.com/splatche3>.

1 Introduction

The simulation of evolutionary scenarios using computer programs is based on the application of combined mathematical models defined by a series of parameters. It is a powerful tool for understanding the mechanisms that shape biodiversity. Compared to deterministic mathematical models, computer simulations have the advantage of considering stochastic processes that better mimic the evolution of species (e.g. Klopstein, et al., 2006). In particular, they allow hypothesis testing by comparing empirical data to those simulated under models representing such hypotheses. When combined with approximate

Bayesian computation (ABC) approaches, they also allow selecting among alternative evolutionary scenarios and estimating parameters of those scenarios (Bertorelle, et al., 2010; Csillery, et al., 2010). They can also be used for comparing and evaluating different analytical tools (e.g. Lopes, et al., 2014; Westesson and Holmes, 2009).

The spatial and temporal dynamics of populations, such as movements in space over time, demographic variation and interactions among populations, plays an important role in evolution. The genetic consequences of those various factors are difficult to disentangle and spatially explicit simulators have been developed for understanding the effects of these combined processes. By spatially explicit, we mean that the simulation takes into account the geographic position of populations.

To the best of our knowledge, the programs SPLATCHE (Currat, et al., 2004) and EASYPOP (Balloux, 2001) were the first spatially explicit computer simulators of genetic diversity available to the scientific community. Other spatially explicit simulation programs pre-dating SPLATCHE were not distributed (Barbujani, et al., 1995; Rendine, et al., 1986). It was initially designed to study the impact of spatial and ecological information on molecular diversity and it modeled migration as movements on a two dimensional stepping-stone lattice (Kimura, 1953). In a few words, SPLATCHE consists in a forward demographic simulation of population demography and migration, followed by a backward coalescent simulation step. In the coalescent step, the ancestry of a sample of gene lineages taken from one or several populations is simulated until the most recent common ancestor of these lineages. Then, genetic diversity of the sample is generated by adding mutations over the simulated coalescent tree. SPLATCHE can handle spatial and environmental complexity through the use of population carrying capacities (e.g. linked to available environmental parameters), migration rates (i.e. directional gene flow) and frictions (i.e. dispersal constraints in different environments) based on user-specified raster maps that can change over time. SPLATCHE and its sequel SPLATCHE2 (Ray, et al., 2010) have been used in many studies, i.e. those on human origin (Ray, et al., 2005) and evolution (Currat and Excoffier, 2005; Lao, et al., 2013; Mona, et al., 2013), on the genetic effects of past range shifts (Banks, et al., 2010; Currat, et al., 2006; Francois, et al., 2010; Gehara, et al., 2013; Ray, et al., 2003; Reid, et al., 2018; Schneider, et al., 2010; Sjodin and Francois, 2011) and climate changes (Brown and Knowles, 2012; Brown, et al., 2016), on admixture between species (Currat, et al., 2008; Durand, et al., 2009) including human and Neanderthals (Currat and Excoffier, 2004; Currat and Excoffier, 2011), as well as on other diverse species (Francois, et al., 2008; White, et al., 2013).

Since the publication of SPLATCHE2 in 2010, new capabilities have been implemented following suggestions from users and authors to provide more realistic simulations and to mimic additional evolutionary scenarios. As a result, we present here a new version of the program that brings together several innovative developments.

2 New capabilities implemented in SPLATCHE3

2.1 Long-distance dispersal

In its previous version, the program could only simulate migration between neighbouring demes under a stepping-stone migration model. SPLATCHE3 implements three new demographic models allowing the simulation of long-distance dispersal (LDD) events. The three LDD models differ by the arrival positions of the LDD events that can be set to target (i) any deme (occupied or empty), (ii) empty demes only, or (iii) occupied demes only. Other user-defined parameters are the proportion of LDD events, the shape of the Gamma distribution used to sample the distance of an LDD event, and the maximum distance of an LDD event. The effects of these models and associated parameters on genetic diversity during range expansion have been explored in Ray and Excoffier (2010). They have also been used to explore the human colonization of Eurasia (Alves, et al., 2016) and the Americas (Branco, et al., 2018).

2.2 Interspecific hybridization

SPLATCHE3 can simulate two interacting populations occupying the same area. Each population is modelled as a layer of interconnected demes spread over space and interactions between populations

(competition and/or admixture) occur between demes located at the same position in their respective layers. The model of admixture implemented in the previous version of the program was criticized on the ground that introgression was assumed as asymmetric (Zhang, 2014). To address this point, we have implemented an alternative model of admixture that is more realistic to simulate hybridization between species. Indeed, the previous model was originally designed to represent the assimilation of individuals from one population to the other one, which affected the size of both populations. In the new model, gene flow is simulated in a more symmetric fashion and it does not modify population densities. We renamed the previous admixture model as “assimilation” and the new one as “hybridization”. Three variants for each model are provided in SPLATCHE3: (i) without inter-specific competition, (ii) with density-dependent inter-specific competition, and (iii) with density-independent inter-specific competition. The new hybridization model was described in detail in Excoffier, et al. (2014) and used to study hybridization between wildcats and domestic cats in Nussberger, et al. (2018).

2.3 Heterogeneous migration rates in space and time

In the previous version of the program, a single migration rate could be assigned to each population layer. In SPLATCHE3, it is now possible to assign different migration rates toward each direction (anisotropic migration) and individually for each deme. It is also possible to vary these individual migration rates over time. This feature has been used to simulate population range contractions caused by glacial periods (Arenas, et al., 2013; Arenas, et al., 2012; Mona, et al., 2014).

2.4 Full disappearance of a population within a deme

A reset of the population size when the carrying capacity of a deme is set to 0 was implemented in SPLATCHE3, allowing simulation of the full disappearance of a population (e.g. due to lack of resources caused by a climatic change). This capability was used in Arenas, et al. (2013); Arenas, et al. (2012); and Mona, et al. (2014).

2.5 Genetic data in serial sampling

Since the first successful retrieval of molecular material from ancient remains (Higuchi, et al., 1984; Paabo, 1985), the production of ancient DNA (aDNA) has increased at an almost exponential pace (e.g. Leonardi, et al., 2017) mainly thanks to rapid advances in sequencing techniques and bioinformatics data processing (e.g. Orlando, et al., 2015). SPLATCHE3 can now simulate genetic data for serial samples through the implementation of the serial coalescent algorithm (Rodrigo and Felsenstein, 1999). This capability of SPLATCHE3 may be used to study ancient DNA and has already been used in two studies investigating population continuity through time (Silva, et al., 2017; Silva, et al., 2018). If one wants to add aDNA post-mortem features, scripts (e.g. bash or R) need to be developed and applied to SPLATCHE3 DNA sequences outputted in ARLEQUIN format (Excoffier and Lischer, 2010).

2.6 Multiple mutation models for DNA evolution

The previous version of the program assumes a single rate of change among nucleotide states when simulating DNA sequence evolution (Jukes and Cantor, 1969). However, more complex patterns of change among nucleotide states are observed in empirical data (Arbiza, et al., 2011) and can be considered to provide more realistic evolutionary analysis (Lemmon and Moriarty, 2004). SPLATCHE3 implements models where 4 nucleotide frequencies and the 6 relative rates of change

among nucleotide states are taken into account. These parameters allow the specification of the most commonly used DNA models of evolution, from the simple JC (Jukes and Cantor, 1969) to the complex GTR (Tavaré, 1986).

2.7 MacOs version

SPLATCHE3 includes a console executable for running on Mac OS, which was not available in previous versions of the program.

3 Conclusion

SPLATCHE3 simulates genetic data under a large combination of environmental and demographic features not available in other spatially explicit simulators. Its main strength is the combination of flexibility and efficiency (through the use of the backward coalescent algorithm), which allows huge gains in computing time when compared to alternative forward individual-based simulators (see Arenas, 2012 and references therein). In this respect, the great flexibility of forward individual-based simulations comes with a cost in computing power and time as they simulate all the individuals of all populations, whereas only a small number of individuals and populations are usually analysed. Forward simulations combined with the coalescent framework are much more efficient because they only simulate genetic information for the sample and its ancestral lineages. This approach is implemented in all versions of the program SPLATCHE, including SPLATCHE3. Its implementation in C++ offers a faster computation speed than the alternative program PHYLOGEOSIM coded in JAVA (Dellicour, et al., 2014). A simple model of range expansion during 1,000 generations from a single deme in a layer composed of 25 to 1,600 demes (Carrying capacity = 300 to 1,000, one DNA sequence of 1,000 bp) results in running-times 2 to 40 faster with SPLATCHE3 than with PHYLOGEOSIM using a 3.6GHz CPU station running on Linux Ubuntu. Other implementations of the coalescent, such as FASTSIMCOAL (Excoffier, et al., 2013; Excoffier and Foll, 2011) or ms (Hudson, 2002), potentially allow performing spatially explicit simulations by specifying an arbitrary migration matrix between all pairs of populations, but for models with complex geographic features the migration matrix becomes huge and is difficult to set up.

SPLATCHE3 is a flexible simulator allowing to investigate a large variety of evolutionary scenarios in a reasonable computational time. The simulation of full genomes would be too time consuming to explore complex evolutionary scenario with SPLATCHE3. However, it is possible to simulate thousands or tens of thousands of independent or partly linked loci and use them as a proxy for full genome data through the computation of summary statistics (e.g. Currat and Excoffier, 2011; Silva, et al., 2018). Running time is mostly affected by the number of simulated generations, number of demes and number of loci. In addition, applying recombination and long-distance dispersal requires additional computational time and memory. SPLATCHE3 is able to integrate various sources of information, such as population genetics, demography, migration, environment and molecular evolution. It allows one to study the effects of population dynamics on the genetic diversity of populations by considering detailed and realistic elements of the landscape that can vary over time. Elements such as continental contours, rivers and mountains can be set to prevent migrations (entirely or partly), while environmental factors (e.g. vegetation types, coastline, deserts) are typically linked to different carrying capacities of demes. The inclusion in SPLATCHE3 of long-distance dispersal, as well as of

refined models of hybridization and mutation opens up new possibilities of research in evolutionary and conservation genetics, especially when it uses the direct testimony of past genetic diversity offered by aDNA.

Acknowledgements

We would like to thank the following people for their help in the development of SPLATCHE3: Nuno Silva, Thomas Goeury, Jérémy Rio, Nicolas Broccard, Stephan Peischl, Isabel Alves, Jason L Brown and David Roessli.

Funding

This work was supported by the Swiss National Science Foundation [31003A-156853 to MC; 31003A_182577 to MC] and by the Spanish Government [RYC-2015-18241 to MA; ED431F 2018/08 to MA].

Conflict of Interest: none declared.

References

- Alves, I., et al. Long-Distance Dispersal Shaped Patterns of Human Genetic Diversity in Eurasia. *Mol Biol Evol* 2016;33(4):946-958.
- Arbiza, L., et al. Genome-Wide Heterogeneity of Nucleotide Substitution Model Fit. *Genome Biol Evol* 2011;3:896-908.
- Arenas, M. Simulation of Molecular Data under Diverse Evolutionary Scenarios. *PLoS Comput Biol* 2012;8(5).
- Arenas, M., et al. Influence of admixture and paleolithic range contractions on current European diversity gradients. *Mol Biol Evol* 2013;30(1):57-61.
- Arenas, M., et al. Consequences of range contractions and range shifts on molecular diversity. *Mol Biol Evol* 2012;29(1):207-218.
- Balloux, F. EASYPOP (version 1.7): A computer program for population genetics simulations. *J Hered* 2001;92(3):301-302.
- Banks, S.C., et al. Genetic structure of a recent climate change-driven range extension. *Mol Ecol* 2010;19(10):2011-2024.
- Barbujani, G., Sokal, R.R. and Oden, N.L. Indo-European origins: a computer-simulation test of five hypotheses. *Am J Phys Anthropol* 1995;96(2):109-132.
- Bertorelle, G., Benazzo, A. and Mona, S. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol Ecol* 2010;19(13):2609-2625.
- Branco, C., et al. Consequences of diverse evolutionary processes on american genetic gradients of modern humans. *Heredity* 2018;121(6):548-556.
- Brown, J.L. and Knowles, L.L. Spatially explicit models of dynamic histories: examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Mol Ecol* 2012;21(15):3757-3775.
- Brown, J.L., et al. Predicting the genetic consequences of future climate change: The power of coupling spatial demography, the coalescent, and historical landscape changes. *Am J Bot* 2016;103(1):153-163.
- Csillery, K., et al. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol* 2010;25(7):410-418.
- Currat, M. and Excoffier, L. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol* 2004;2(12):2264-2274.
- Currat, M. and Excoffier, L. The effect of the Neolithic expansion on European molecular diversity. *Proc Biol Sci* 2005;272(1564):679-688.
- Currat, M. and Excoffier, L. Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *P Natl Acad Sci USA* 2011;108(37):15129-15134.
- Currat, M., et al. 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* 2006;313(5784):172; author reply 172.
- Currat, M., Ray, N. and Excoffier, L. SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Mol Ecol Notes* 2004;4(1):139-142.
- Currat, M., et al. The hidden side of invasions: massive introgression by local genes. *Evolution; international journal of organic evolution* 2008;62(8):1908-1920.
- Dellicour, S., et al. Comparing Phylogeographic Hypotheses by Simulating DNA

- Sequences under a Spatially Explicit Model of Coalescence. *Molecular Biology and Evolution* 2014;31(12):3359-3372.
- Durand, E., et al. Spatial Inference of Admixture Proportions and Secondary Contact Zones. *Molecular Biology and Evolution* 2009;26(9):1963-1973.
- Excoffier, L., et al. Robust Demographic Inference from Genomic and SNP Data. *Plos Genet* 2013;9(10).
- Excoffier, L. and Foll, M. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 2011;27(9):1332-1334.
- Excoffier, L. and Lischer, H.E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010;10(3):564-567.
- Excoffier, L., Quilodran, C.S. and Currat, M. Models of hybridization during range expansions and their application to recent human evolution. In: MV, D.A.S., editor, *Cultural Developments in the Eurasian Paleolithic and the Origin of Anatomically Modern Humans*. Dept. of the Institute of Archaeology and Ethnography: SB RAS; 2014. p. 122-137.
- Francois, O., et al. Demographic history of European populations of *Arabidopsis thaliana*. *Plos Genet* 2008;4(5).
- Francois, O., et al. Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol* 2010;27(6):1257-1268.
- Gehara, M., Summers, K. and Brown, J.L. Population expansion, isolation and selection: novel insights on the evolution of color diversity in the strawberry poison frog. *Evol Ecol* 2013;27(4):797-824.
- Higuchi, R., et al. DNA-Sequences from the Quagga, an Extinct Member of the Horse Family. *Nature* 1984;312(5991):282-284.
- Hudson, R.R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 2002;18(2):337-338.
- Jukes, T. and Cantor, C. Evolution of protein molecules. In: Munro, H.N., editor, *Mamalian Protein Metabolism*. New York: Academic press; 1969. p. 21-132.
- Kimura, M. "Stepping-stone" model of population. *Annual Report of National Institute of Genetics* 1953;3:62-63.
- Klopfstein, S., Currat, M. and Excoffier, L. The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* 2006;23(3):482-490.
- Lao, O., et al. Clinal distribution of human genomic diversity across the Netherlands despite archaeological evidence for genetic discontinuities in Dutch population history. *Investig Genet* 2013;4(1):9.
- Lemmon, A.R. and Moriarty, E.C. The importance of proper model assumption in Bayesian phylogenetics. *Syst Biol* 2004;53(2):265-277.
- Leonardi, M., et al. Evolutionary Patterns and Processes: Lessons from Ancient DNA. *Syst Biol* 2017;66(1):E1-E29.
- Lopes, J.S., et al. Coestimation of recombination, substitution and molecular adaptation rates by approximate Bayesian computation. *Heredity* 2014;112(3):255-264.
- Mona, S., et al. Investigating sex-specific dynamics using uniparental markers: West New Guinea as a case study. *Ecol Evol* 2013;3(8):2647-2660.
- Mona, S., et al. Genetic consequences of habitat fragmentation during a range expansion. *Heredity* 2014;112(3):291-299.
- Nussberger, B., et al. Range expansion as an explanation for introgression in European wildcats. *Biol Conserv* 2018;218:49-56.
- Orlando, L., Gilbert, M.T.P. and Willerslev, E. APPLICATIONS OF NEXT-GENERATION SEQUENCING Reconstructing ancient genomes and epigenomes. *Nat Rev Genet* 2015;16(7):395-408.
- Paabo, S. Molecular-Cloning of Ancient Egyptian Mummy DNA. *Nature* 1985;314(6012):644-645.
- Ray, N., et al. Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Research* 2005;15(8):1161-1167.
- Ray, N., Currat, M. and Excoffier, L. Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution* 2003;20(1):76-86.
- Ray, N., et al. SPLATCHE2: a spatially-explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* 2010.
- Ray, N. and Excoffier, L. A first step towards inferring levels of long-distance dispersal during past expansions. *Mol Ecol Resour* 2010;10(5):902-914.
- Reid, B.N., et al. Disentangling the genetic effects of refugial isolation and range expansion in a trans-continentally distributed species. *Heredity (Edinb)* 2018.
- Rendine, S., Piazza, A. and Cavalli-Sforza, L. Simulation and separation by principal components of multiple demic expansions in Europe. *Am. Nat.* 1986;128(5):681-706.
- Rodrigo, A.G. and Felsenstein, J. Coalescent approaches to HIV-1 population genetics. In: A., C.K., editor, *The evolution of HIV*. Baltimore, Maryland: John Hopkins Univ. Press; 1999. p. 233-272.
- Schneider, N., et al. Signals of recent spatial expansions in the grey mouse lemur (*Microcebus murinus*). *Bmc Evol Biol* 2010;10.
- Silva, N.M., Rio, J. and Currat, M. Investigating population continuity with ancient DNA under a spatially explicit simulation framework. *Bmc Genet* 2017;18.
- Silva, N.M., et al. Bayesian estimation of partial population continuity using ancient DNA and spatially explicit simulations. *Evol Appl* 2018;11(9):1642-1655.
- Sjodin, P. and Francois, O. Wave-of-advance models of the diffusion of the Y chromosome haplogroup R1b1b2 in Europe. *PLoS One* 2011;6(6):e21592.
- Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura, R.M., editor, *Some mathematical questions in biology - DNA sequence analysis*. Amer Math Soc; 1986. p. 57-86.
- Westesson, O. and Holmes, I. Accurate Detection of Recombinant Breakpoints in Whole-Genome Alignments. *Plos Comput Biol* 2009;5(3).
- White, T.A., et al. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Mol Ecol* 2013;22(11):2971-2985.
- Zhang, D.Y. Demographic model of admixture predicts symmetric introgression when a species expands into the range of another: A comment on Currat et al. (2008). *J Syst Evol* 2014;52(1):35-39.