

PROGRAM NOTE

SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity

M. CURRAT,*†‡ N. RAY*†‡ and L. EXCOFFIER*

*Computational and Molecular Population Genetics Laboratory, Zoological Institute, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland, †Genetics and Biometry Laboratory, Anthropology and Ecology Department, University of Geneva, Rue Gustave Revilliod 12, 1227 Carouge, Switzerland

Abstract

We present a program called **SPLATCHE** (**S**patial **A**nd **T**emporal **C**oalescences in **H**eterogeneous **E**nvironments) to simulate the molecular diversity of samples of genes in an environmentally heterogeneous world. Simulations are performed by, first, simulating the colonization of the world using environmental information to constrain migrations and local densities. These simulated densities and migration rates recorded over time and space are then used to simulate genetic diversity under a coalescent framework. The program thus virtually allows the translation of ecological information into molecular diversity, a novel approach that can be used to study the effect of climatic change on genetic diversity.

Keywords: coalescent simulation, demography, heterogeneous environment, molecular diversity, spatial expansion

Received 24 September 2003; revision accepted 20 November 2003

It is now widely admitted that climatic changes had (and still have) a profound impact on the distribution and the genetic diversity of many animal and plant species (e.g. Taberlet *et al.* 1998; Hewitt 2000; Barnes *et al.* 2002; Walther *et al.* 2002). At any given time, the distribution of a species and its diversity are heavily influenced by the environmental heterogeneity (e.g. Brachet *et al.* 1999; Wakeley & Aliacar 2001; Hanski & Ovaskainen 2003). For example, spatial fluctuations in food resources can lead to major differences in the sizes of subpopulations, and some landscape elements can act as barriers to migrations or in the contrary facilitate migration and act as corridors along which a larger amount of individuals can migrate. Therefore, due to the importance of the environment on real genetic processes, we have developed a program called **SPLATCHE** for the simulation of sampled molecular diversity, explicitly taking into account the spatial heterogeneity of the environment. **SPLATCHE** is actually based on a two-step simulation process: environmental information is first used to simulate (forward in time) the demographic and migration history of a set of subpopulations, and the resulting information is

then used in a coalescent framework to generate (backward in time) the genealogy and the diversity of genes sampled at one or several locations (see Fig. 1). These two phases are enclosed in two separate modules of the program.

The simulation of the first demographic phase occurs in a two-dimensional stepping-stone (2DSS, Kimura & Weiss 1964), which is defined as an array of regularly spaced subpopulations or demes. Each deme can exchange migrants with its four neighbours. Information on local environments [such as vegetation or topographic maps, which can be imported from a Geographical Information System (GIS) package], is translated into two characteristics attached to each deme: a carrying capacity K and a friction value F . K represents here the maximum number of genes (or haploid individuals) that can be sustained by local resources, and F expresses the (relative) difficulty in moving through a deme, using relative values ranging from 0 (lowest friction, no barrier to migration) to 1 (highest friction, complete barrier to migration). Arbitrary levels of environmental heterogeneity can be considered. For instance, user-defined maps can be completely homogeneous, in term of K and F , but they can also reflect a more realistic world taking into account particular landscape features such as rivers, deserts, or mountains. These features can then be assigned specific F -values to reflect their potential roles as corridors

Correspondence: Professor Laurent Excoffier. Fax: + 41 31 631 48 88; E-mail: laurent.excoffier@zoo.unibe.ch

†These two authors have contributed equally to this work.

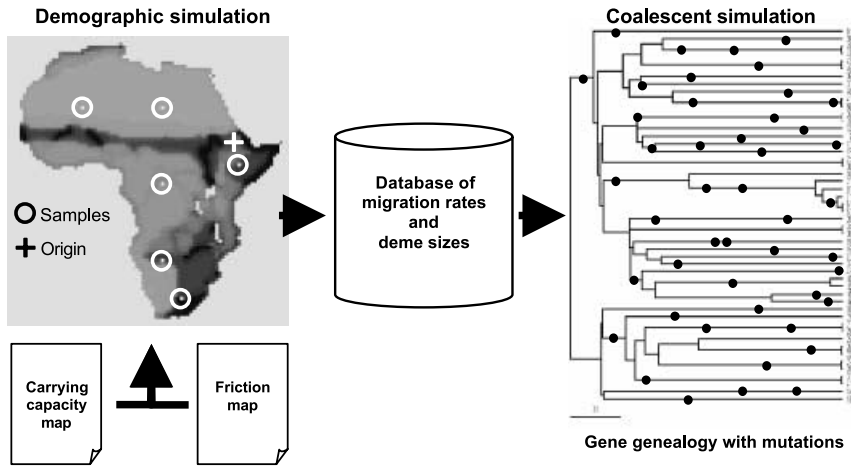


Fig. 1 Schematic view of the two-step simulation processes (demography and coalescent) necessary to generate genetic diversity at a given locus. The left pane illustrate the fact that forward demographic simulations are carried out from spatial information on potential carrying capacities and friction in different environment. The resulting simulated deme densities and immigration rates are used to perform coalescent simulations (right pane). Mutations following a given model are then sprayed on the coalescent tree to generate molecular diversity, and are shown as black dots on the gene genealogy.

or barriers to migration. Different K -values can also be assigned to reveal spatial differences in resource allocation. Different maps of K and F can be used at different times to simulate environmental changes. The maps to use and the time at which these changes occur are simply listed in input files. The format for the input maps is *ascii raster*, which can be generated by most GIS packages. A typical demographic simulation starts at one arbitrarily chosen deme, whose size and geographical location can be defined by the user in an input file. In each occupied deme, a growth phase is followed by a migration phase. The assumed growth model within deme is a standard logistic growth characterized by a user-defined intrinsic rate of growth per generation r , assumed constant over deme and over time, and a carrying capacity K depending on the environment. During the migration phase, the number of emigrants for every deme is computed as $N_t m$, where N_t is the local deme size at time t , and m is the migration rate. The number of emigrants can be allowed to vary stochastically as a Poisson variable with mean $N_t m$. The number of emigrants sent to each neighbouring deme is chosen from a multinomial distribution, with directional probabilities inversely proportional to the relative frictions of the neighbouring demes (see Ray 2003 for details). The sending of migrants to unoccupied demes results typically in a wave of advance of the whole population as shown on Fig. 2. The velocity and the shape of the edge of this wave depend on various parameters of the model, such as F , K , m , and r (see Ray 2003).

A modified version of the SIMCOAL program (Excoffier *et al.* 2000) has been integrated into SPLATCHE, and it is used to generate the molecular diversity of one or several samples. All former specifications of SIMCOAL are available except the 'historical events' formerly used to resize deme sizes and migration patterns between demes. Now, instead of manually setting the sizes of the demes and the pattern of migrations prevailing between demes at different time,

this information is obtained from the previous demographic simulation round (see Fig. 1). The population size N_t of a given deme at time t is used to compute the probability of coalescence for every pair of genes found in that deme at time t . The number of emigrants (forward in time) from deme i to deme j at time t is used to compute the backward probability (backward in time) for a gene in deme j to migrate into deme i . SPLATCHE can generate restriction fragment length polymorphism, short tandem repeat (microsatellite data), DNA sequence data, and mere allele frequencies. The number and the spatial location of genes that must be simulated are specified in an input file.

In addition to text output in the ARLEQUIN format (Schneider *et al.* 2000), several graphical outputs are also available after a demographic simulation: deme size and number of migrants received from neighbouring demes through time, times of first colonization. Carrying capacities and frictions assigned to each deme can also be visualized as grayscale or colour bitmaps. They can alternatively be saved as *ascii raster* files to be imported into any GIS program. An innovative feature of SPLATCHE is the visualization of the spatial components of the coalescence trees, which can be translated into three different kinds of bitmaps: (i) the spatial distribution of the density of coalescence events, as shown for example in Ray *et al.* (2003); (ii) the genealogical connection between the nodes of the tree, in order to visualize the coalescent tree spatially onto the simulated world; (iii) the spatial distribution of the most recent common ancestor of all genes (MRCA), either for the whole tree or independently for each population sample. Two additional output files also allow one to obtain, for each sample, the spatial distribution of all coalescent events, and the time to the MRCA (T_{MRCA}).

The ability to translate environmental information into molecular diversity, via a demographic model, allows for a wide array of applications. Using a very simple homogeneous world, it has been possible to study the effect of a

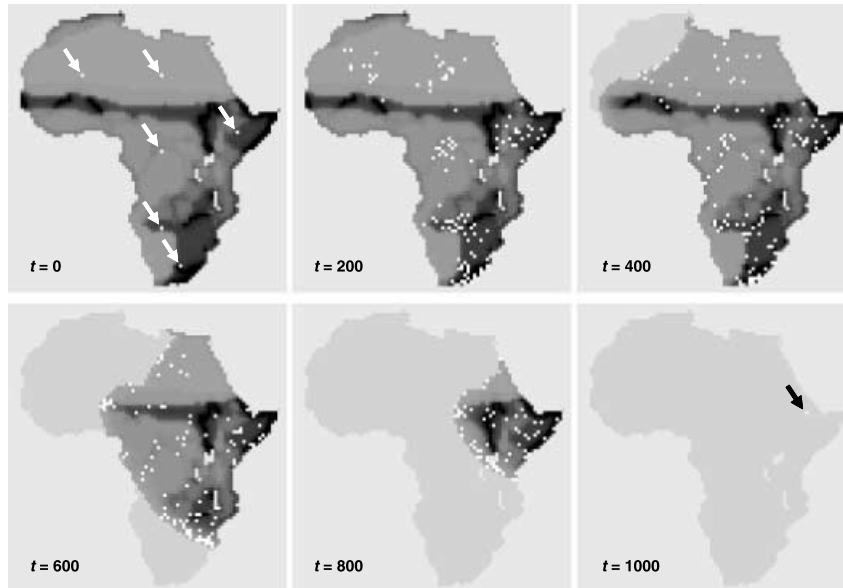


Fig. 2 A graphical example of a coalescent simulation in a hypothetical case assuming that Africa has been colonized from a single deme located in Eastern Africa (as shown by the black arrow on the lower right pane) 1000 generations ago. White arrows in the upper left pane indicate the locations of the sample demes at time zero, each containing 30 genes. The time t is expressed in number of generations from present, and goes backward in time. Darker demes have higher densities. At all times, white demes are those in which there is at least one gene lineage whose descendants can be traced forward to sampled genes. One sees that (going backward in time) gene lineages initially located in the sampled demes migrate to neighbouring demes and progressively diffuse in the environment. They are brought back to the origin of the settlement process by the shrinking of the occupied portion of the simulated world, which corresponds to the inverse of the assumed range expansion having started from East Africa.

range expansion on molecular diversity while varying parameters such as the carrying capacity and migration rates (Ray *et al.* 2003). Theoretical expectation of many statistics can be obtained with this simulation framework, because thousands of coalescent simulations can be generated in a reasonable amount of time, and later analysed with ARLEQUIN. The role of particular environmental or demographic variables can then be thoroughly assessed by analysing the molecular diversity, and obtaining the variance of the genetic processes. By using heterogeneous carrying capacity and friction maps, it is possible to consider very realistic situations in SPLATCHE. Specific metapopulation models could be implemented and SPLATCHE could also be used to make prediction about the spatial range and molecular diversity of a set of population after a change in the environment (man-mediated or following a climatic change). The graphical possibilities are also a major advantage of SPLATCHE. The visualized superposition of the demographic and coalescent processes can lead to a better understanding of some phenomenon, for instance in the study of the effect of migration corridor or spatial bottlenecks on genetic diversity, or the discovery of spatial regions that show an excess of migrants or of coalescent events. Finally, the educational value of SPLATCHE should be underlined, as many insights can be gained by visualizing the movements of genes through space and time in a realistic environment.

Executable Windows version of SPLATCHE, user guide and example files can be downloaded from <http://cmpg.unibe.ch/software/splatche>. Depending on the size of the simulated world and the number of generations, large amounts of RAM are required to guarantee reasonable execution time. For instance, a simulation of 10 000 demes over 4000 generations requires about 400 Mb of free RAM.

Acknowledgements

We are grateful to Stefan Schneider and Pierre Berthier for their computing assistance. This work was supported by a Swiss NSF grant no. 31-054059-98 to LE.

References

- Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science*, **295**, 2267–2270.
- Brachet S, Olivier I, Godelle B *et al.* (1999) Dispersal and metapopulation viability in a heterogeneous landscape. *Journal of Theoretical Biology*, **198**, 479–495.
- Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, **91**, 506–510.
- Hanski I, Ovaskainen O (2003) Metapopulation theory for fragmented landscapes. *Theoretical Population Biology*, **64**, 119–127.

- Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
- Kimura M, Weiss WH (1964) The stepping stone model of genetic structure and the decrease of genetic correlation with distance. *Genetics*, **49**, 561–576.
- Ray N (2003) *Modélisation de la démographie des populations humaines préhistoriques à l'aide de données environnementales et génétiques*. PhD Thesis, University of Geneva, Switzerland.
- Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**, 76–86.
- Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN: a software for population genetics data analysis*. User manual. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Wakeley J, Aliacar N (2001) Gene genealogies in a metapopulation. *Genetics*, **159**, 893–905.
- Walther GR, Post E, Convey P *et al.* (2002) Ecological responses to recent climate change. *Nature*, **416**, 389–395.