SPLATCHE2

USER MANUAL

2010

Version: 22.09.2010

Authors:

N. Ray[1], M. Currat[2], M. Foll[3,4], L. Excoffier[3,4]

1- EnviroSPACE laboratory, Climate Change and Climate Impacts, Institute for Environmental Sciences, University of Geneva, Switzerland.

2- Laboratory of anthropology, genetics and peopling history, Department of Anthropology, University of Geneva, Switzerland.

3- Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Berne, Switzerland

4- Swiss Institute of Bioinformatics, Switzerland

with the help of D. Wegmann, S. Klopfstein, G. Laval,
S. Neuenschwander, G. Hamilton, S. Mona, A. Sramkova

Table of contents

Table of Figures

1 Introduction

The goal of this user manual is to describe the technical aspects of the current version of the SPLATCHE2 program (for SPatiaL And Temporal Coalescences in Heterogeneous Environments, version 2.0). SPLATCHE2 is derived from the SPLATCHE program that has been released in 2004. This coalescent-based program has been designed to model complex demographies, such as range expansions, in a heterogeneous and varying environment, and to simulate the resulting molecular diversity of samples of genes. Its goal is to allow using realistic environmental models 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 (Ray, Currat et al. 2005; Foll and Gaggiotti 2006) and especially the genetic consequences of range expansion (Ray, Currat et al. 2003; Hamilton, Currat et al. 2005; Currat, Excoffier et al. 2006; Klopfstein, Currat et al. 2006; Wegmann, Currat et al. 2006).

Many improvements and new functionalities have been added to the original version of SPLATCHE (see next section), allowing the simulation of a much wider range of evolutionary scenarios. Some of these new features have already been applied independently, such as the simulation of admixture and competition between populations (Currat and Excoffier 2004; Currat and Excoffier 2005; Currat, Ruedi et al. 2008; Francois, Currat et al. 2010) and the use of multiple source populations (Ray, Currat et al. 2005). Those extensions have now been merged in SPLATCHE2, together with the possibility to simulate linked or partially linked loci, to allow for multiple coalescent events per generation, and to generate Single Nucleotide Polymorphism (SNPs). Moreover, SPLATCHE2 is also available as a console version aimed to facilitate its use on computer clusters.

A new, more flexible, settings file now allows one to have full control over all input settings, which greatly facilitates running SPLATCHE2 in console mode. These settings can also be set through the command line. The settings are described in Section 5.2.

1.1 Changes from SPLATCHE version 1.1

Numerous changes have occurred since the published version SPLATCHE 1.1. Changes of some internal software components have allowed an overall speed increase and to improve program stability. We list below the new functionalities of SPLATCHE2:

- A new format for the **input settings file**
- Simulation of **competition** and **admixture** between two coexisting populations
- Simulation of unlinked, partially linked or fully linked loci, using a **recombination** model

- Generation of Single Nucleotide Polymorphisms (**SNPs**)
- **Multiple coalescent events** per generation
- **Environmental fluctuations over time** (also possible with two populations)
- **Multiple source** populations
- **Console** version (Windows and Linux) in addition to the graphical version

1.2 Citation and download

Nicolas Ray, Mathias Currat, Matthieu Foll and Laurent Excoffier: SPLATCHE2: a spatially-explicit simulation framework for complex demography, genetic admixture and recombination. Bioinformatics (Submitted)

The last version of SPLATCHE2 can be freely downloaded from the web site http://www.splatche.com

2 General approach

SPLATCHE2 simulations are based on a two-phase process, namely the forward and the backward (coalescent) steps. The first phase is the simulation, forward in time, of a two dimensional stepping stone (Kimura and Weiss 1964) world, which is an array of equally spaced demes (or sub populations). Each deme exchanges migrants with its four neighbours depending on three parameters: the carrying capacity (the number of individuals that can be sustained by local resources within a deme), the migration rate (the per-generation probability of an individual to move out of a deme) and the friction (the relative ease to move into a deme).

Both carrying capacities and frictions may be kept homogeneous over the whole simulated array, or they may be set to different values for each deme. This latter possibility is done by providing specific maps of corresponding values over the landscape. In this way, it is possible to take into account realistic geographic and environmental features, by varying carrying capacities of the demes according to the quality of the habitat and change friction values in order to model rivers, mountains or roads. Notably, such maps can be held fixed during the whole simulation process or may change after a specified number of generations to take into account historical events, such as recent human induced fragmentation of the habitat (Quemere, Crouau-Roy et al. 2010), cycle ages with progressive reduction of available landscape (Taberlet, Fumagalli et al. 1998; Hewitt 2004) and more.

2.1 Forward demographic simulations

The forward simulation starts from one (or more) user defined deme(s), which send(s) migrants to neighbouring demes (Figure 1). Migrations to empty demes represent new colonization events. The density of occupied demes is logistically regulated, implying a local demographic expansion of newly colonized demes. Hence, at each generation, a logistic regulation step (regulated by the carrying capacity and the growth rate) is followed by a migration step (regulated by the effective number of individuals, the migration rate and the friction). The whole simulation process, which lasts for a user-defined number of generations, results in a wave of advance of the whole population, with a shape and at a speed determined by the demographic parameters and the friction values.

One important feature of SPLATCHE2 is that migrations occur from one deme to a neighboring one, but also in the reverse direction and especially that migrations do not stop when a deme is fully colonized. It means that migrations continue even at demographic equilibrium (when all the simulated world has been colonized).

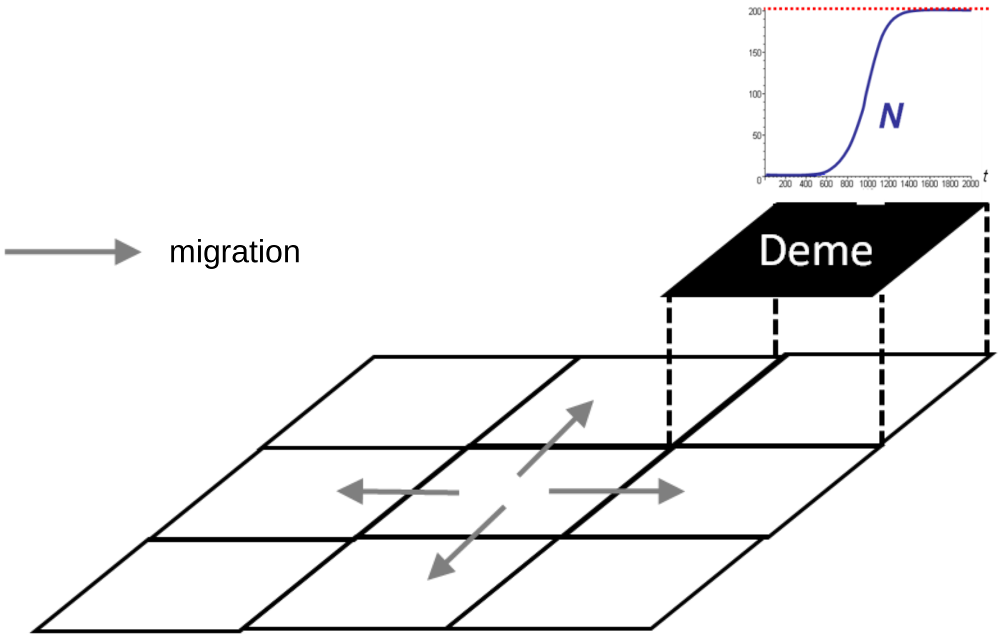


Figure 1 Schematic view of an array of demes where the simulations occur

2.2 Backward genetic simulations

Each generation, the demographic and migration histories of every deme are stored in a database in computer memory (not on disk). This database is then used in the backward coalescent step (Hudson 1990) of the SPLATCHE2 program. This second phase of the algorithm now starts from the present generation, and proceeds backward in time (Figure 2). The effective number of individuals present in a deme is used to compute the probability of a coalescent event, and the migration rates and friction values determine the probability of each sampled genes to emigrate, backward in time, to the surrounding demes. The coalescent process stops after all genes have coalesced, most often in the deme where the range expansion started. If more than one expansion origins (source populations) have been set, additional parameters defining the migration rates among the origins prior to the expansion need to be specified.

2.3 Sampling

Genes can be sampled in any number and in any demes of the simulated world, and therefore at different spatial scales (within deme, within nearby demes -a patch-, or at very remote locations). The genetic markers simulated can be outputted in text files in ARLEQUIN/ARLSUMSTAT format (Excoffier and Lischer 2010) so that the genetic diversity can be monitored at any of such scales level.

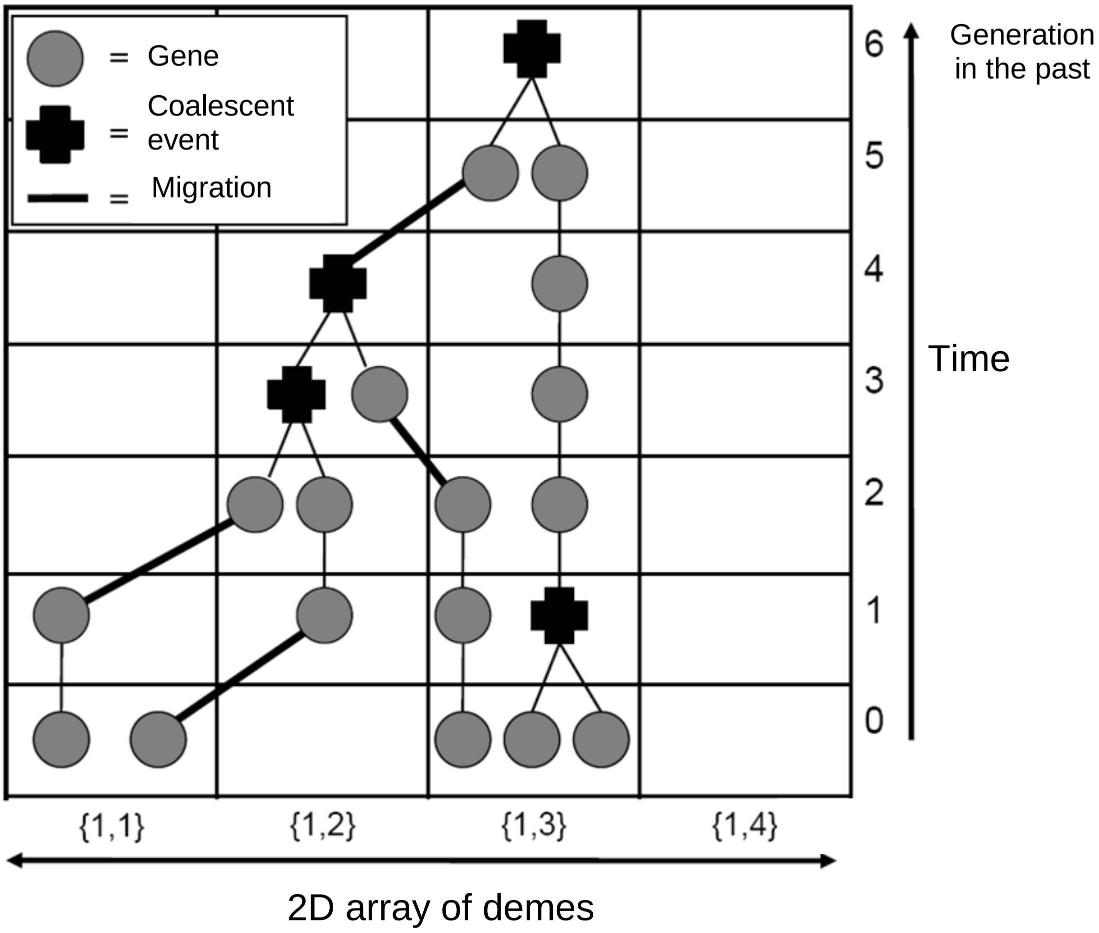


Figure 2 Schematic view of the backward coalescent process

2.4 *Simulation of two coexisting and interacting populations*

Two co-existing and interacting populations can be simulated using SPLATCHE2. Technically, two layers of demes representing the virtual world are superimposed (Figure 3). Each layer represents one population. The two populations can coexist independently (without any interaction) in each layer but they can also be in competition for the resources and exchange genes by interbreeding.

Source populations can be located in different places in each layer but there are two main constraints: First, there must be at least one source for the population A at the start of the simulation (generation 0); Second, individuals constituting the source(s) of population B are taken from the corresponding deme in population A (e.g. the deme located at the same place in the layer A). This means that population B is created by a group of individuals belonging initially to population A. It can represent, for example, a speciation event.

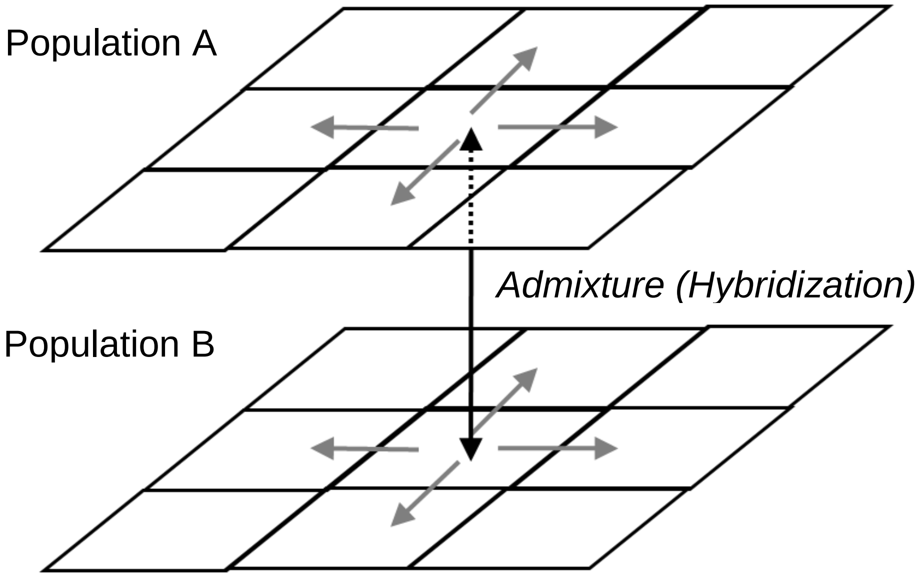


Figure 3 Schematic representation of the two superimposed layers of demes

For instance, Figure 4 shows a schematic representation of two successive spatial and demographic expansions. Each population expansion takes place in one layer and gene flow (admixture) can occur between the two populations. The demographic parameters of each population can be set independently. It is also possible to simulate competition between the two populations, potentially leading to the extinction of one population.

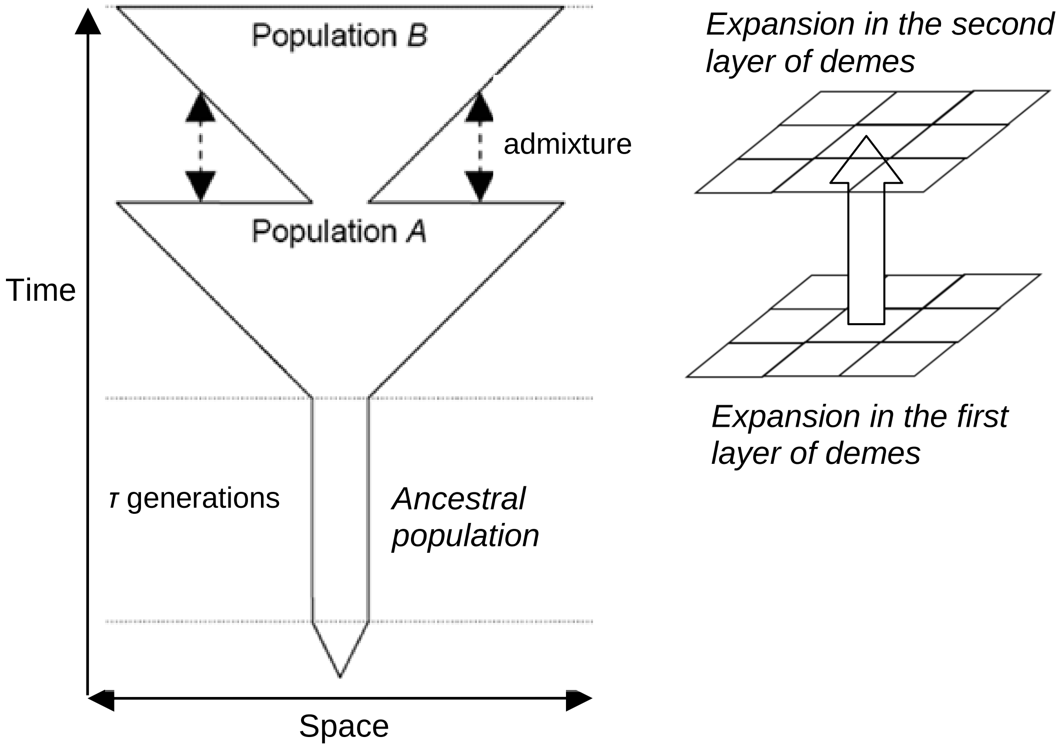

Population B
admixture
Population A
Time
τ generations
Ancestral population
Space
Expansion in the second layer of demes
Expansion in the first layer of demes

Figure 4 Schematic representation of the two successive population expansions with admixture

2.5 Computer requirement

SPLATCHE2 does not need any minimal computer requirement, and should run on any descent modern desktop. However, because demographic information for every deme is stored in computer memory, the complexity of a simulation is limited by the characteristics of the computer. RAM memory requirement is roughly proportional to the product of the number of demes by the number of generations. For example, using the setting file *settings_test_1layer.txt* distributed with SPLATCHE2, 400 MB of memory are necessary to simulate an expansion in 1700 active demes during 1,000 generations. Roughly 1 GB of memory is necessary to simulate 5,000 demes during 10,000 generations. One way to decrease memory requirement by almost 50% is to set the parameter *AllowShortIntForNumberOfIndividuals* to 1 (see chapter 5.2). In that case, the storage of the deme densities requires less memory, but the drawback is that these densities must not exceed 16,000 individuals per deme at any generation during the simulation.

In order for the user to have an idea about time requirement when performing simulations with SPLATCHE2, the table below shows the time it takes for various simulations on two different computers and operating systems. We used one of the datasets given as example in the web site (*settings_test_1layer.txt*, which is constituted by 1700 active demes). We performed one demographic simulation and two genetic simulations: one with 400 haploid individuals constituted by DNA sequences of 500 pb length and a second one with 100 independent STRs in 200 diploids individuals. Note that those times have been computed without recombination. When simulating recombination, the time needed for a genetic simulation increases proportionally with the recombination rate. Also note that the requirement in RAM memory increases with the recombination rate and that the program may ultimately crash if this rate is too high related to the other parameters (mainly sequence length, carrying capacity *K* and migration rate *m*).

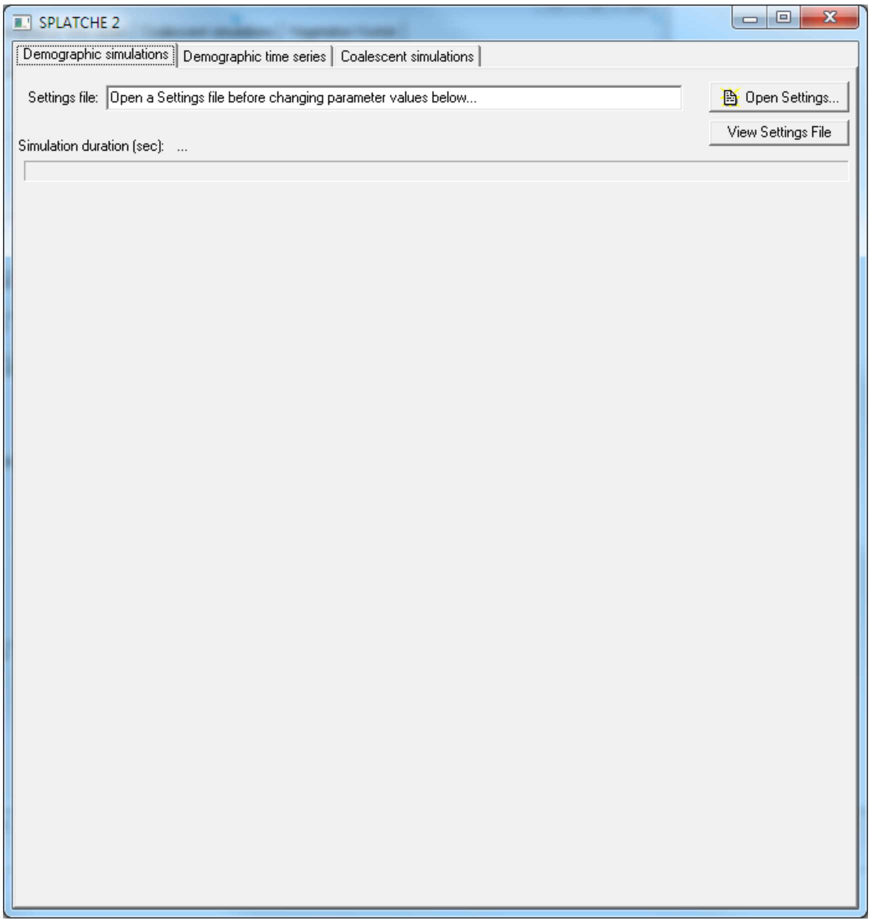
Dataset	**Graphical version on Windows XP (2.66GHz CPU)**	**Console version on Linux 64 bits (2.3GHz CPU)**
settings_test_1layer.txt *500 generations*	Demography: 9 sec Genetics (DNA 500pb): 1 sec Genetics (100 STRs): 35 sec.	Demography: 9 sec Genetics (DNA 500pb): <1 sec Genetics (100 STRs): 6 sec.
settings_test_1layer.txt *1000 generations*	Demography: 20 sec Genetics (DNA 500pb): 2 sec Genetics (100 STRs): 55 sec.	Demography: 14 sec Genetics (DNA 500pb): 1 sec Genetics (100 STRs): 10 sec.

3 Overview of the interface

3.1 Windows graphical interface

The Windows graphical interface allows the user to interactively modify some of the input data (e.g., carrying capacity, frictions, geographic locations of source populations and sampled populations) and to explore the output results (e.g., range expansion, coalescence process). Over the years, the authors of SPLATCHE2 have found the graphical visualization of the simulations extremely important to (*i*) check the validity of input data, (*ii*) understand the influence of input settings on demographic and genetic outputs, (*iii*) discover patterns and behaviors of expanding populations and their associated coalescent process.

When opened, the graphical interface looks as follows:

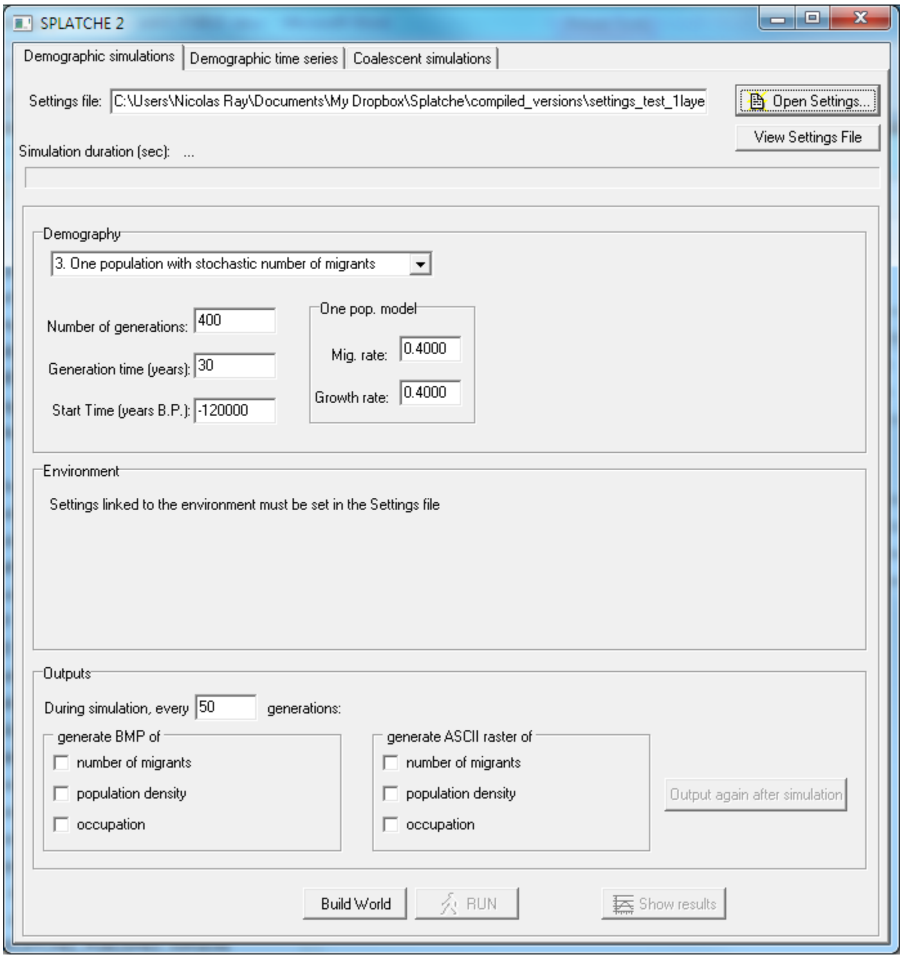


Settings file: Open a Settings file before changing parameter values below...

Simulation duration (sec): ...

Figure 5 SPLATCHE2 opening panel

The first step is to open a settings file, which will fill the corresponding settings fields in the graphical interface. Note that only a restricted number of parameters may be changed using the graphical interface, most of them can only be changed in the setting files.



Settings file: C:\Users\Nicolas Ray\Documents\My Dropbox\Splatche\compiled_versions\settings_test_1laye

Simulation duration (sec): ...

Demography

3. One population with stochastic number of migrants

Number of generations: 400

Generation time (years): 30

Start Time (years B.P.): -120000

One pop. model

Mig. rate: 0.4000

Growth rate: 0.4000

Environment

Settings linked to the environment must be set in the Settings file

Outputs

During simulation, every 50 generations:

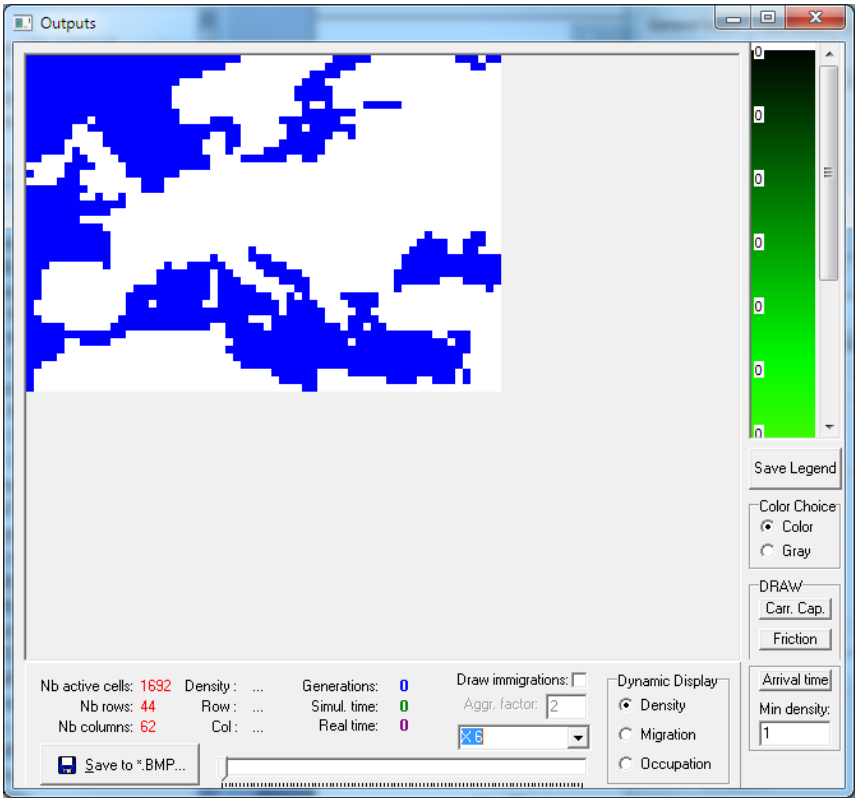
generate BMP of
- □ number of migrants
- □ population density
- □ occupation

generate ASCII raster of
- □ number of migrants
- □ population density
- □ occupation

Figure 6 *Demographic simulation* panel

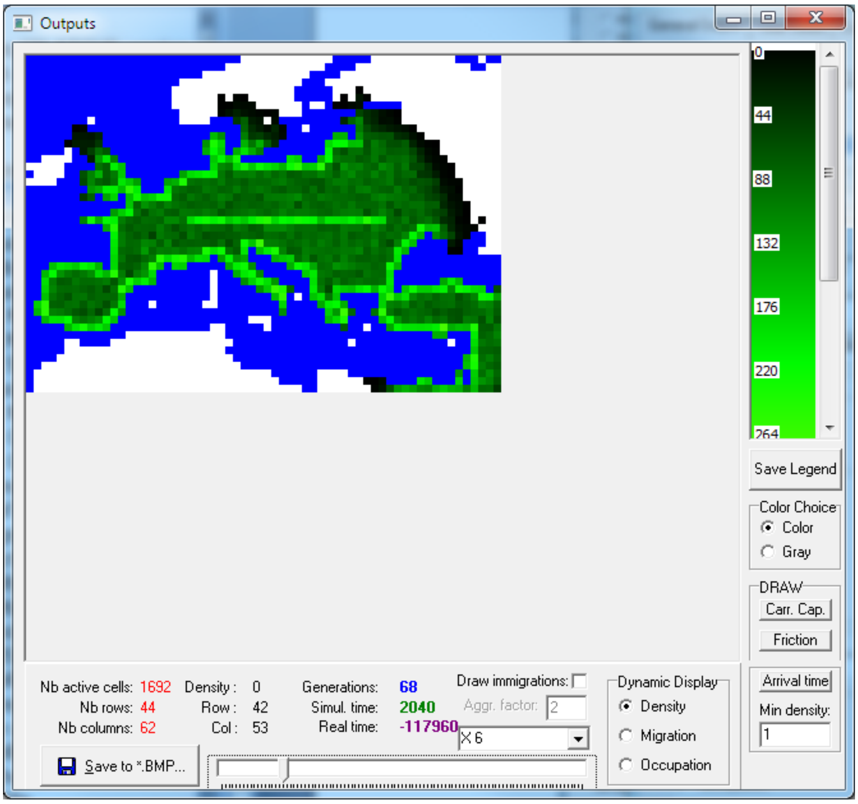
After the successful reading of a settings file, three panels are made available, but only the first one (*Demographic simulations*, (Figure 6) is needed at this stage. Some settings can be changed through the interface (e.g. identification of the migration scenario), while others (e.g. carrying capacities, frictions) must be changed directly in the settings files.

Once all simulation settings values have been chosen, the user must click on the "*Build World*" button. This button opens a new window (called *Outputs*, Figure 7), instantiates the world (i.e. the virtual 2D array of demes), allocates appropriate computer memory for the simulations, and shows the instantiated world (with sea or *NoData* demes in blue, see Figure 7).



Nb active cells: 1692
Nb rows: 44
Nb columns: 62
Density : ...
Row : ...
Col : ...
Generations: 0
Simul. time: 0
Real time: 0
Draw immigrations: □
Aggr. factor: 2
X 6
Dynamic Display: ● Density
Color Choice: ● Color
Min density: 1

Figure 7 *Outputs* windows after the “build world” step and before the forward simulation step



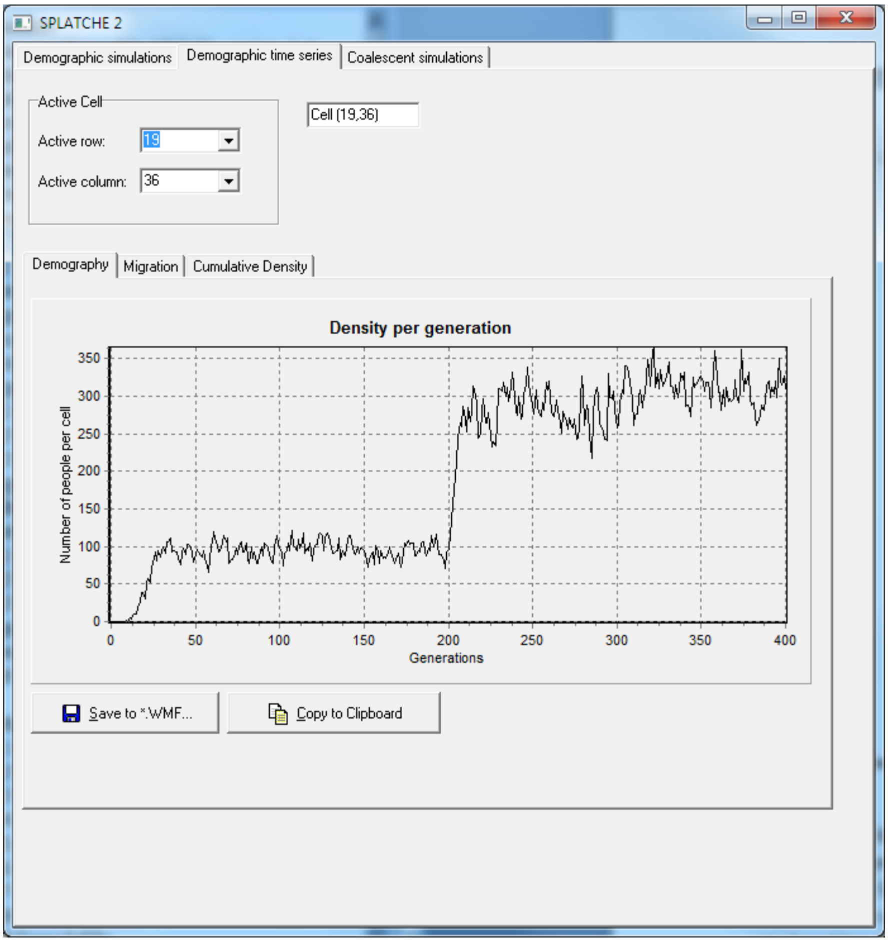
Nb active cells: 1692
Nb rows: 44
Nb columns: 62
Density : 0
Row : 42
Col : 53
Generations: 68
Simul. time: 2040
Real time: -117960
Draw immigrations: □
Aggr. factor: 2
X 6
Dynamic Display: ● Density
Color Choice: ● Color
Min density: 1

Figure 8 *Outputs* windows: exploration of the demographic history using the sliding bar

Once the user is satisfied with input data and the selected settings, the next step is to click on the *Run* button (now available) on the *Demographic simulations* window (Figure 6). At the end of the simulations the output window shows the end state of the demographic simulation. Note that these simulations can take some time depending on the migration model and on the number of generations, but the computation progress is indicated on the *Simulation duration* bar (Figure 6). The demographic history can then be explored using the sliding bar at the bottom of the window (see Figure 8).

By clicking on a deme on the resulting world bitmap, one can access its full demographic history that appears in the second panel (*Demographic time series*, Figure 9) of the main windows. The *Demographic time series* panel shows the demographic history of any simulated cell defined by its geographic coordinates.

The first sub-panel called "*Demography*" shows the evolution of the density within the deme, generation after generation. Two other sub-panels are also available ("*Migration*" and "*Cumulative Density*") that allows one to visualize the migration histories and the cumulative total population size, respectively.



Active Cell

Active row: 19

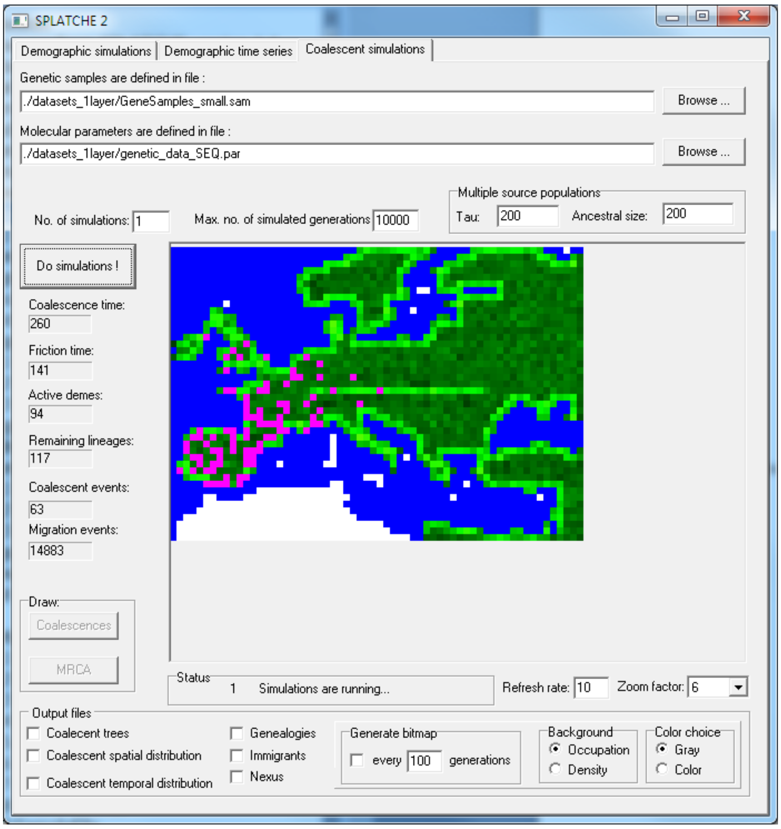
Active column: 36

Cell (19,36)

Figure 9 *Demographic time series*: exploration of the demographic history of a given deme

Finally, the third panel (*Coalescent simulations*) on the main window allows one to launch the genetic simulations. This panel is filled with parameters values found in the Settings file, and, again, only a few of them (number of simulations and parameters related to the multiple source populations) can be changed in the graphical interface before launching the simulation, which is done by clicking on the "*Do simulations!*" button. During the coalescent backward simulation, the demes occupied by at least one gene are depicted as violet dots.

Note that one independent coalescent simulation is done per independent marker. At the end of the genetic simulations, bitmaps showing the locations of the coalescent events and the locations of the MRCA can be opened by clicking on the "*Coalescences*" and "*MRCA*" buttons, respectively.



Genetic samples are defined in file : ./datasets_1layer/GeneSamples_small.sam

Molecular parameters are defined in file : ./datasets_1layer/genetic_data_SEQ.par

No. of simulations: 1

Max. no. of simulated generations 10000

Multiple source populations: Tau: 200 Ancestral size: 200

Coalescence time: 260

Friction time: 141

Active demes: 94

Remaining lineages: 117

Coalescent events: 63

Migration events: 14883

Status: 1 Simulations are running...

Refresh rate: 10 Zoom factor: 6

Output files

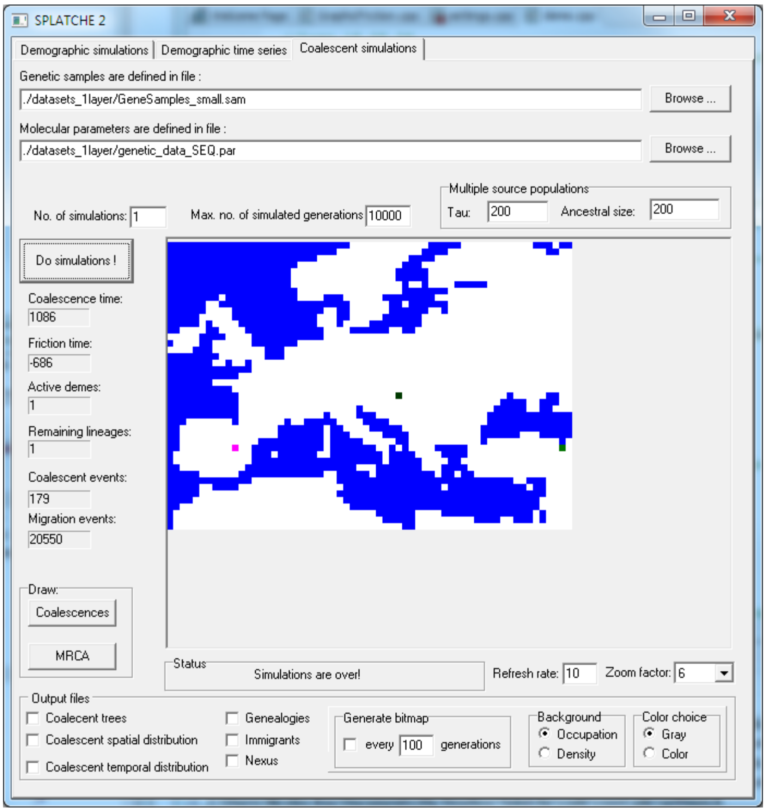
□ Coalecent trees
□ Coalescent spatial distribution
□ Coalescent temporal distribution
□ Genealogies
□ Immigrants
□ Nexus

Generate bitmap: □ every 100 generations

Background: ● Occupation ○ Density

Color choice: ● Gray ○ Color

Figure 10 *Coalescent simulations* panel showing the moving genes over the simulated world



Genetic samples are defined in file : ./datasets_1layer/GeneSamples_small.sam

Molecular parameters are defined in file : ./datasets_1layer/genetic_data_SEQ.par

No. of simulations: 1

Max. no. of simulated generations 10000

Multiple source populations: Tau: 200 Ancestral size: 200

Coalescence time: 1086

Friction time: -686

Active demes: 1

Remaining lineages: 1

Coalescent events: 179

Migration events: 20550

Status: Simulations are over!

Refresh rate: 10

Zoom factor: 6

Output files:
☐ Coalecent trees
☐ Coalescent spatial distribution
☐ Coalescent temporal distribution
☐ Genealogies
☐ Immigrants
☐ Nexus

Generate bitmap: ☐ every 100 generations

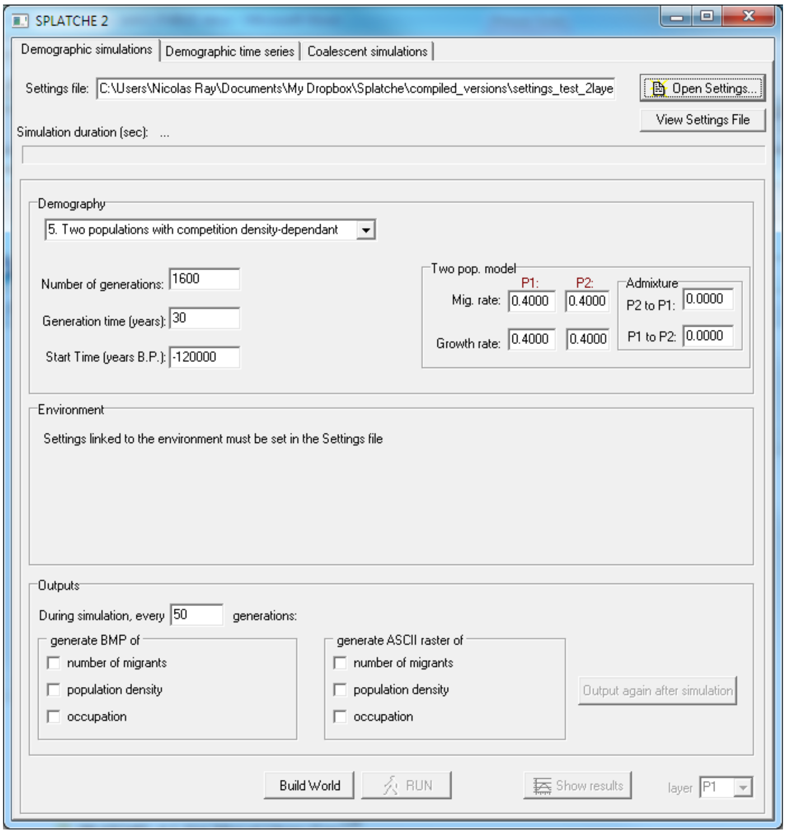
Background: ● Occupation ○ Density

Color choice: ● Gray ○ Color

Figure 11 ***Coalescent simulations*** **panel: end of a genetic simulation**

3.1.1 Windows graphical interface for two populations

When opening a setting file defining two populations (DoublePopulationMode=1), a series of new fields appear on the graphical version (Figure 12). The acronyms P1 and P2 stand for “Population one” and “Population two”. Admixture rates between P1 and P2 may be asymmetrical and set independently (see chapter 5.2). The “layer” dropdown menu on the bottom right of the Demographic simulations panel indicates which population (P1 or P2) is displayed in the *Demographic time series* panel, *Coalescent simulations* panel, and the Outputs window.



Settings file: C:\Users\Nicolas Ray\Documents\My Dropbox\Splatche\compiled_versions\settings_test_2laye

Simulation duration (sec): ...

5. Two populations with competition density-dependant

Number of generations: 1600

Generation time (years): 30

Start Time (years B.P.): -120000

	P1:	P2:
Mig. rate:	0.4000	0.4000
Growth rate:	0.4000	0.4000

P2 to P1: 0.0000

P1 to P2: 0.0000

Settings linked to the environment must be set in the Settings file

During simulation, every 50 generations:

generate BMP of
- □ number of migrants
- □ population density
- □ occupation

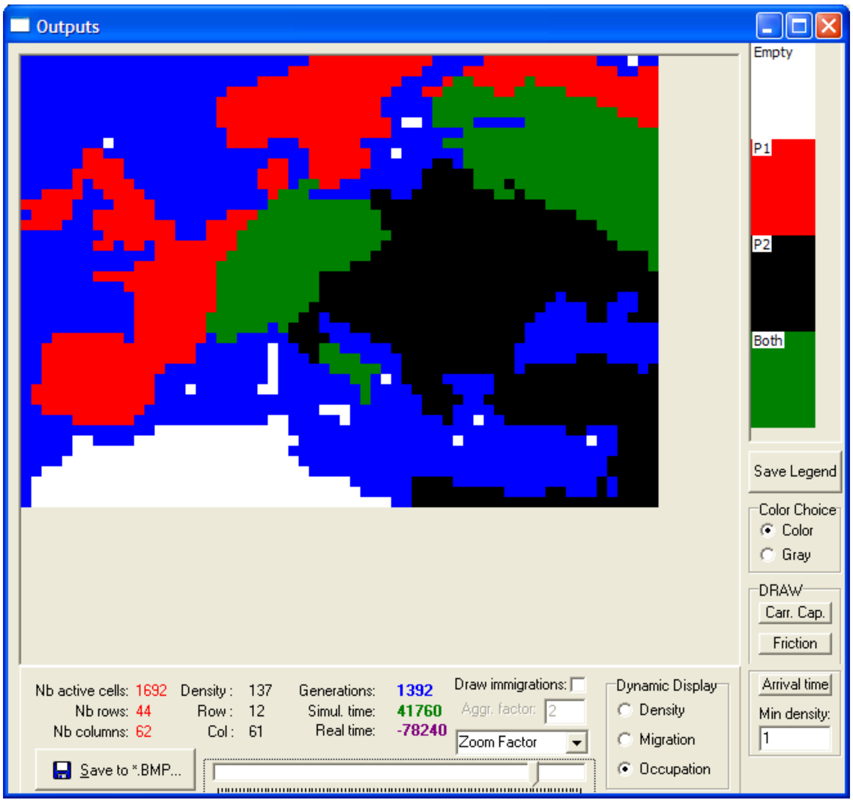
generate ASCII raster of
- □ number of migrants
- □ population density
- □ occupation

layer P1

Figure 12 ***Demographic simulations*** **panel for two populations**

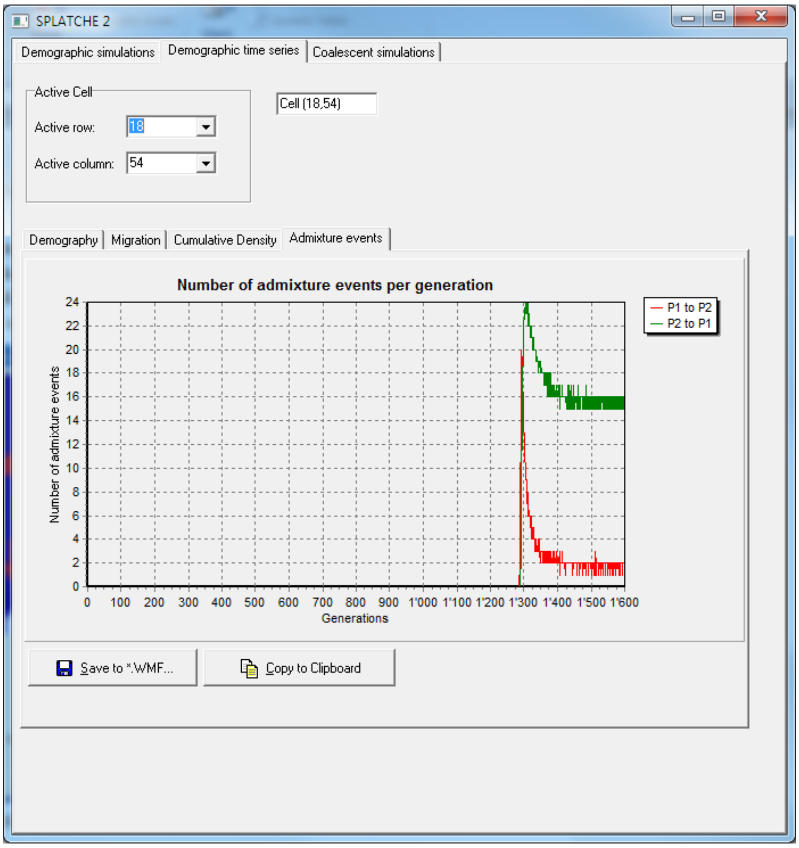
In the outputs window, after a demographic run, there is a new option available called “occupation” (see Figure 13). This option displays in different colors the demes that are occupied either by one or two populations at any time of the simulation.

After a demographic run, there is an additional sub-panel called “Admixture events” available in the *Demographic time series* panel (Figure 14). This panel displays the number of realized admixture events (or introgressions) in both directions (e.g. gene flow from P1 to P2 or the reverse) during the simulation and for any given deme.



Nb active cells: 1692
Nb rows: 44
Nb columns: 62
Density : 137
Row : 12
Col : 61
Generations: 1392
Simul. time: 41760
Real time: -78240
Draw immigrations: □
Aggr. factor: 2
Zoom Factor
Dynamic Display
○ Density
○ Migration
● Occupation
Color Choice
● Color
○ Gray
Min density: 1

Figure 13 *Outputs* window showing the occupation of the world by the two populations



Active Cell
Active row: 18
Active column: 54
Cell (18,54)

Figure 14 *Admixture events* subpanel

3.2 Console interface (Windows and Linux)

The console versions (Windows and Linux) have been created to make it easier to launch parallel instances of SPLATCHE2 on computer clusters. With these versions, launching a set of simulations (demographic + genetic) is simple, because it only involves calling the program at the command prompt with the name of the settings file, for example under Linux:

```
$ splatche2_lin_64 settings_test_1layer.txt
```

Output messages will appear during the simulations, for example:

```
$ splatche2_lin_64 settings_test_1layer.txt
Reading inputfile 'settings_test_1layer.txt' .....
Random generator initialized with seed=-1437263
Building World ... OK
Simulating from 'settings_test_1layer.txt'
Simulating demography, iteration no. 1...
OK
Simulating genetics, iteration no. 1...
OK
ALL SIMULATIONS AND PROGRAMS TERMINATED SUCCESFULLY!
```

The Linux version is available under 32-bits or 64-bits version of the OS.

4 Demographic models

The demographic simulation regulates local population growth and migration. Three migration models are available when using one layer. With two layers, three models can be chosen based on the level of competition between the two populations. We describe these models hereafter.

4.1 Local population growth

The local logistic growth is the same for one or two layers. The logistic population growth of each deme follows a standard logistic curve, of the form

$$N_{t+1} = N_t \left(1 + r \frac{K - N_t}{K} \right),$$

where K is the deme-specific carrying capacity, and r is the growth rate. Note that the sequence of the demographic steps at each generation step is as follows: (1) demographic growth, (2) local migration.

4.2 Friction

In SPLATCHE2, the friction attributed to a focal deme controls the ease/difficulty of immigrating to that focal deme from neighboring demes. When the number of emigrants is computed in a focal deme, the migration model will "look" at the frictions in the neighboring demes and will distribute the emigrants in the neighboring demes according to their relative frictions.

Frictions (*f*) are expressed as relative numbers and must be set up as 0<*f*<1, with friction close to 0 expressing ease of movement and frictions close to 1 expressing difficulty of movements. If *f*≤0 or *f*≥1, the deme is considered isolated and no migrant will reach it.

When computing the directional percentage of emigrants (p_i), or the directional probability of emigrations, the following formula is applied:

$$p_i = \frac{1}{f_j \sum_{j=1}^{n} \frac{1}{f_j}}$$

where f_j is the friction of the deme in direction *j* (north, south, east or west) over the *n* available neighboring demes, and $\sum p_i = 1$.

Note that the total number of emigrants from a focal deme is always spread among the *n* available neighboring demes, even if only 3, 2, or 1 of such neighboring deme(s) are available (i.e. there is no loss of emigrants, or in other terms there is no absorbing boundaries around demes).

4.3 Migration models with one population

For the migration part of the demography, three models are available in SPLATCHE2. To use any of these models, the ID (1, 2, or 3, as defined below) of the chosen model must be indicated for the settings *ChosenDemographicModel* in the Settings file (see Section 5.2).

4.3.1 Model 1: Migration model with even number of emigrants

The number of emigrants M_i in any of the *n* available directions is computed, for each generation, as $M_i = floor(p_i \cdot m \cdot N_t)$, where m is the migration rate, N_t is the population density of the deme at generation t, p_i is the directional percentage of emigrants, and *floor* means that the fractional part of the number is truncated. The total number of emigrants *M* from a deme is the sum of the M_i, and is therefore always a multiple of the available *n* neighbors.

The speed of this model has been improved compared to the corresponding one in SPLATCHE ver. 1.1. The fractional parts of M_i at time *t* (prior to applying the *floor* function) are kept in a variable and are accounted for in the next generation by adding them to the deme density.

4.3.2 Model 2. Migration model with absolute number of emigrants

Same as Model 1, but the fractional part of M_i is not truncated. Instead, a multinomial distribution is used to split M emigrants to the neighboring demes (see Ray 2003), according to the directional probability of emigration (p_i). This ensures that there are always M emigrants that are sent. The drawback of this technique is that it requires the drawing of random numbers, which increases simulation time.

4.3.3 Model 3. Stochastic migration model with absolute number of emigrants

Same as Model 2, but the number of emigrants M varies stochastically as a Poisson variable centered around $N_t m$. This model is the one used in Ray *et al.* (2005).

4.4 Migration model with two populations

For all three demographic models available with two layers (*DoublePopulationMode* = 1), the migration model is equivalent to the model 1 described above (section 4.3.1). It means that no migration stochasticity is available with two layers, in all cases an even number of emigrants is distributed in the neighboring demes. With two populations, the life cycle of a population at a given generation is as follows: admixture (see section 4.5), logistic regulation incorporating competition (see section 4.6), followed by migration. This life cycle thus assumes that migration is at the adult stage.

4.5 Admixture model with two populations

The frequency of admixture events is assumed to be density-dependent. Within a given deme, each of the N_i individuals from the *i*-th population has a probability

$$A_{ij} = \gamma_{ij}(2N_i N_j)/(N_i + N_j)^2$$

to reproduce successfully with one of the N_j members of the *j*-th population, and γ_{ij} represents the probability that such a mating results in a fertile offspring.Alternatively, γ_{ij} could represent the relative fitness of hybrid individuals or an index of disassortative mating. γ_{ij} and γ_{jj} are defined by the user in setting the parameters *MigrRate_P1_to_P2* and *MigrRate_P1_to_P2* respectively. Following admixture, population densities are then first updated as

$$N_i^{t+1} = N_i^t(1 - A_{ij}) + A_{ij} N_j^t$$

4.6 Competition models with two populations

The three following models are only available when the *DoublePopulationMode* parameter has been set to one.

4.6.1 Model 4: No competition

In that case, there is no competition between the two populations. The demography of both population is regulated independently using the model 1 described above (section 4.3.1). This model was used in Currat et al. (2008).

4.6.2 Model 5: Competition density-dependant

This model includes competition density-dependant between the two populations. Our model of density regulation incorporating competition is based on the Lotka–Volterra interspecific competition model, which is itself an extension of the logistic growth model (Volterra 1926; Lotka 1932). For each population i, a new density N_i^{t+1} is calculated from the former density N_i^t as

$$N_i^{t+1} = N_i^t(1 + r_i(K_i - N_i^t - \alpha_{ij}N_j^t)/K_j)$$

Where r_i is the intrinsic growth rate, K_i the carrying capacity, N_j is the density in the other population from the same cell and α_{ij} is an asymmetric competition coefficient. An α_{ij} value of 1 implies that individuals of the j-th population have as much influence on those of population i as on their own conspecific, or that competition between populations is as strong as competition within a population. Lower values of α_{ij} indicate lower levels of competition between populations than within populations; a value of zero implies no competition between individuals from different populations. Model 5 assumes that competition is density-dependant and α_{ij} is thus calculated as

$$\mathrm{\alpha_{ij}} = \mathrm{N_j^t}/(\mathrm{N_i^t} + \mathrm{N_j^t}),$$

reflecting the fact that the influence of the members of a population on the other grows with its density. This model was used in Currat & Excoffier (2004; 2005).

4.6.3 Model 6: Competition density-independent (classical Lotka-Volterra)

This model is the same than model 5 but competition coefficients $\mathrm{\alpha_{ij}}$ are set to 1 and do not depend on the current densities. This model was used in Currat et al. (2008).

5 Settings, input and output files

5.1 Description of the Settings file

The settings file has been completely redesigned compared to SPLATCHE ver. 1.1. For an example, see the sample settings file available on the SPLATCHE website. Each setting is now set through a "*SettingsName*=*value*" pair, for instance:

```
ChosenDemographicModel=3
```

Note that Settings names should not be changed and that they are case sensitive. No space should be put before and after the "=" sign. Any line in the Settings file with the "#" sign in front of it is commented and therefore ignored. Empty lines are also ignored. The order of the lines in the Settings file does not matter, but we advice users to keep the original order for clarity.

Some of the settings are not mandatory and can therefore be commented if one does not want to use them. For mandatory settings, the program will throw an error if these settings are not defined. For some settings, it is only when defined (i.e. not commented) that the corresponding machinery is available. For example, the taking into account of frictions is only available if the setting "FrictionChoice" is defined.

All settings that are paths to other text files must be set either with absolute paths (e.g. C://documents/file.txt) or with relative path (e.g. ./file.txt). The relative paths must be set relative to where the SPLATCHE2 executable is located.

5.2 List of settings with a short definition

Below is a list of all available settings, presented as they appear in the sample Settings file available on the SPLATCHE2 website. Although settings can appear in any order in the Settings file, they have been grouped here by way of their use. When necessary, the reader is sent to a detailed description of the settings or its format in the next Section 5.3.

Setting name	**Definition**
Parameters linked to filenames	
PopDensityFile	Path to a text file with the locations, initial densities, etc. of the initial population(s). See section 5.3.1. (MANDATORY)
PresVegetationFile	Path to an Ascii file describing the initial world (array of demes). See section 5.3.2 (MANDATORY)
HydroFile	Path to an Ascii file describing hydrology (rivers). See section 5.3.2 (NOT MANDATORY)
RoughnessTopoFile	Path to an Ascii file with the roughness values. See section 5.3.2 (MANDATORY only if FrictionChoice=1 or =2)
Veg2KFile	Path to a text file holding paths to tables making the correspondence between vegetation categories and carrying capacity values (MANDATORY). See section 5.3.3
Veg2FFile	Path to a text file holding paths to tables making the correspondence between vegetation categories and friction values (NOT MANDATORY). See section 5.3.3
Parameters linked to demographic simulations	
ChosenDemographicModel	Identification (number) for the demographic model (MANDATORY). See Section 2.5.
EndTime	Number of simulated generations (MANDATORY)
GenerationTime	Duration of a generation (in years), used in several functions needing it (MANDATORY)
GrowthRate	Intrinsic growth rate used in the logistic function (MANDATORY)

MigrationRate	Migration rate for neighboring deme migration (MANDATORY). See section 5.3.4
AllowSourcePopulationOverflow	The value indicates whether the density *N* that is attributed to the initial deme is spread over neighboring demes when *N* is greater than the carrying capacity (NOT MANDATORY). See section 5.3.5
TauValue	Tau value (in generations). Only used with multiple source populations. Backward in time, this is the time between the onset of the expansion and the timing at which all remaining lineages are brought into one small deme (NOT MANDATORY). See section 5.3.6
AncestralSize	Size of the ancestral deme at time Tau (defined in the setting *TauValue*). (NOT MANDATORY). See section 5.3.6
ArrivalCellFile	Text file (*.col) with coordinates of demes for which arrival times are needed (MANDATORY). See section 5.3.7
Parameters linked to physical environment	
FrictionChoice	Choice of friction type (0: vegetation,1: roughness topography,2: both). If friction needs to be taken into account, this parameter must be defined (along with "Veg2FFile" and/or "RoughnessTopoFile"). If not defined, friction is uniform (NOT MANDATORY). See section 5.3.8
RealBPTime	Real time (in years before present (BP)) of the start of the simulation. The number must be a negative integer. This parameter is linked to the graphical display of the real time in any simulation (MANDATORY)
RiverFrictionChangeFactor	This is a factor [>0] that increases or decreases the friction of the river cells (as defined in the HydroFile input file). The friction of these cells is simply multiplied by this factor. In case of "double populations", the change is applied on both friction maps. (NOT MANDATORY)
RiverCarCapChangeFactor	This is a factor [>0] that increases or decreases the carrying capacity of the river cells (as

	defined in the HydroFile input file). The carrying capacity of these cells is simply multiplied by this factor. In case of "double populations", the change is applied on both friction maps. (NOT MANDATORY)
CoastFrictionChangeFactor	This is a factor [>0] that increases or decreases the friction of the coast cells (coasts are automatically defined). The friction of these cells is simply multiplied by this factor. In case of "double populations", the change is applied on both friction maps. (NOT MANDATORY)
CoastCarCapChangeFactor	This is a factor [>0] that increases or decreases the carrying capacity of the coast cells (coasts are automatically defined). The carrying capacity of these cells is simply multiplied by this factor. In case of "double populations", the change is applied on both carrying capacity maps. (NOT MANDATORY)
Parameters linked to genetic simulations	
SampleFile	Text file containing the coordinates and sizes (+other info) of the genetic samples. It must have a "*.sam" extension (MANDATORY). See section 5.3.9.
GeneticFile	Text file (*.par) containing the definition of markers property. This is the main file for genetic settings, with or without recombination (MANDATORY). See Section 5.3.10
NumGeneticSimulations	Number of genetic simulations following the demographic simulation. Each genetic simulation independently reconstructs a genealogy of the sampled genes and produces one ARLEQUIN output file (MANDATORY)
GenotypicData	Choice of genotypic or haplotypic data (1:genotypic; 0: haplotypic)(MANDATORY) Note that haploid data are always simulated, only the way output are represented does change.
MaxNumGenerations	Maximum number of total generations for a genetic simulation. This number corresponds to the number of generation of the demographic

	simulation + the extra generations for the coalescence process prior to time 0. After this time the process stops if the genealogy has not been correctly reconstructed and none output file is produced (MANDATORY)
GenealogiesFile	Generation of genealogy files *.tri (0:no ; 1: yes) (MANDATORY). See section 5.3.13
ImmigDistrFile	Generation of immigrants distribution file *.nm (0:no ; 1: yes) (MANDATORY). See section 5.3.14
Parameters linked to double populations	
DoublePopulationMode	Flag to indicate that double populations (two layers) version is used (0: no ; 1: yes) (MANDATORY)
GrowthRate_P2	Intrinsic growth rate for the second population (P2) used in the logistic function. (MANDATORY IF *DoublePopulationMode*=1)
MigrationRate_P2	Migration rate for the second population (P2) to neighboring deme migration. (MANDATORY IF *DoublePopulationMode*=1)
MigrRate_P1_to_P2	Admixture rate from P1 to P2 [0-1], see section 5.3.15. (MANDATORY IF *DoublePopulationMode*=1)
MigrRate_P2_to_P1	Admixture rate from P2 to P1 [0-1], see section 5.3.15 (MANDATORY IF *DoublePopulationMode*=1)
Veg2K_P2_File	Text file for population 2 (P2) holding the carrying capacity values for each vegetation category. (MANDATORY IF *DoublePopulationMode*=1)
Veg2F_P2_File	Text file holding the friction values for each vegetation category. (NOT MANDATORY)
PropFile	Compute theproportion of sampled genes in the second population (P2) which are descendent from the original source population P2 (0:no ; 1: yes), see section (MANDATORY IF *DoublePopulationMode*=1)

Parameters linked to various other outputs	
GenerateOutputMigrationBMP	Generate output BMP of migrations (0:no ; 1: yes) (MANDATORY)
GenerateOutputMDensityBMP	Generate output BMP of densities (0:no ; 1: yes) (MANDATORY)
GenerateOutputOccupationBMP	Generate output BMP of occupations (0:no ; 1: yes) (MANDATORY)
GenerateOutputMigrationASCII	Generate output ASCII of migrations (0:no ; 1: yes) (MANDATORY)
GenerateOutputMDensityASCII	Generate output ASCII of densities (0:no ; 1: yes) (MANDATORY)
GenerateOutputOccupationASCII	Generate output ASCII of occupations (0:no ; 1: yes) (MANDATORY)
Miscelaneous Parameters	
AllowShortIntForNumberOfIndividuals	Allow "short int" (instead of int) to be used for the database. This reduces by half the amount of RAM necessary. Only do that if you are sure that your numbers of individuals (i.e. population densities and number of migrants) never exceed 16,000 in any single deme! (NOT MANDATORY)

5.3 *Detailed description of settings and outputs*

5.3.1 PopDensityFile

A file, called “dens_init_europe_small.txt” in the example Settings file, is used to specify the place(s) of origin of the simulated population. This file contains a first line with the number of source populations to be considered, and then one line per source population. Each of these lines consists of 10 fields separated by “tab” or “space” character.

Example:

```
3
Paleo         100    12     60     5      10     100    0.4    0      0
Iberic        100    12     10     0      50     50     0.1    0      0
Central       50     20     35     0      150    150    0.2    0      0
```

These 10 fields (parameters) correspond to:

1. The name of the source populations (without space).

2. The size of the source population, in number of effective haploid individuals.

3. & **4**. The geographic coordinates of the source population (latitude and longitude). SPLATCHE2 will determine in which particular deme the coordinates of the population corresponds. Coordinates do not need to be in a particular units (e.g. decimal degrees), but they needs to be in the same units that the coordinates defined in the header of the environmental files (see Section 5.3.2). Each coordinate must also fall within the extent of the world defined by the environmental files and they should not fall in "sea" (NoData) demes.

5. A resize parameter: it is the size of the source population before the beginning of the expansion. This parameter is used only for genetic simulations. If this parameter is set to 0, then the size of the source population before the onset of the expansion is regarded as being equal to the initial size (set in the second column).

6 & 7. The time (6) in number of generations before the beginning of the expansion when a second resize (7) occurs. It could be seen as the duration of the bottleneck and the population size before the bottleneck.

8. The emigration rate from the current source to the other sources before the time of the expansion. This rate is only used in case of multiple source populations (e.g. multiple sources for the first population P1 at generation 0).

9. Index of the layer where the source appears: 0 means the first layer (P1) while 1 indicates the second layer (P2). Note that: *i)* If there is only one layer (population) simulated (*DoublePopulationMode*=0), all sources must start in P1; *ii)* if two layers are simulated (*DoublePopulationMode*=1), one source must start in P2, otherwise the second population will not be created. *iii)* If this parameter is set to 1 (starting in population P2) then a number of individual equal to the size of the source population (column 2) will be taken from P1 at the same location in the array of demes, in order to found the new population P2. It results that if the number of individuals currently belonging to the deme in P1 is smaller or equal to the size of the population source in P2, then all individuals then all individuals in P1 will be taken. It also means that if P1 has not been colonized, then P2 source cannot be created, and the second layer is not colonized.

10. The expansion time. If two populations are simulated (*DoublePopulationMode*=1), the founding of the second population (P2) should start after the beginning of the expansion of the first population (P1). This parameters indicates the time (in number of generation) at which the source must start. Note the following constraints: *ii)* it must be at least one source starting in P1 at generation 0; *ii)* P1 sources can only start at generation 0.

Figure 15 shows a schematic example two parallel expansions in the first layer 1. Two source populations start in the first layer (P1), both at generation 0. Both source populations passed through a bottleneck prior to their expansions and share a common ancestral population tau generations before the onset of their expansions. Compare this scheme to the one represented in Figure 4 that shows another example of two source populations but in two different layers. One source for the first population (P1) at generation 0 and one source for the second population (P2) a given number of generations later.

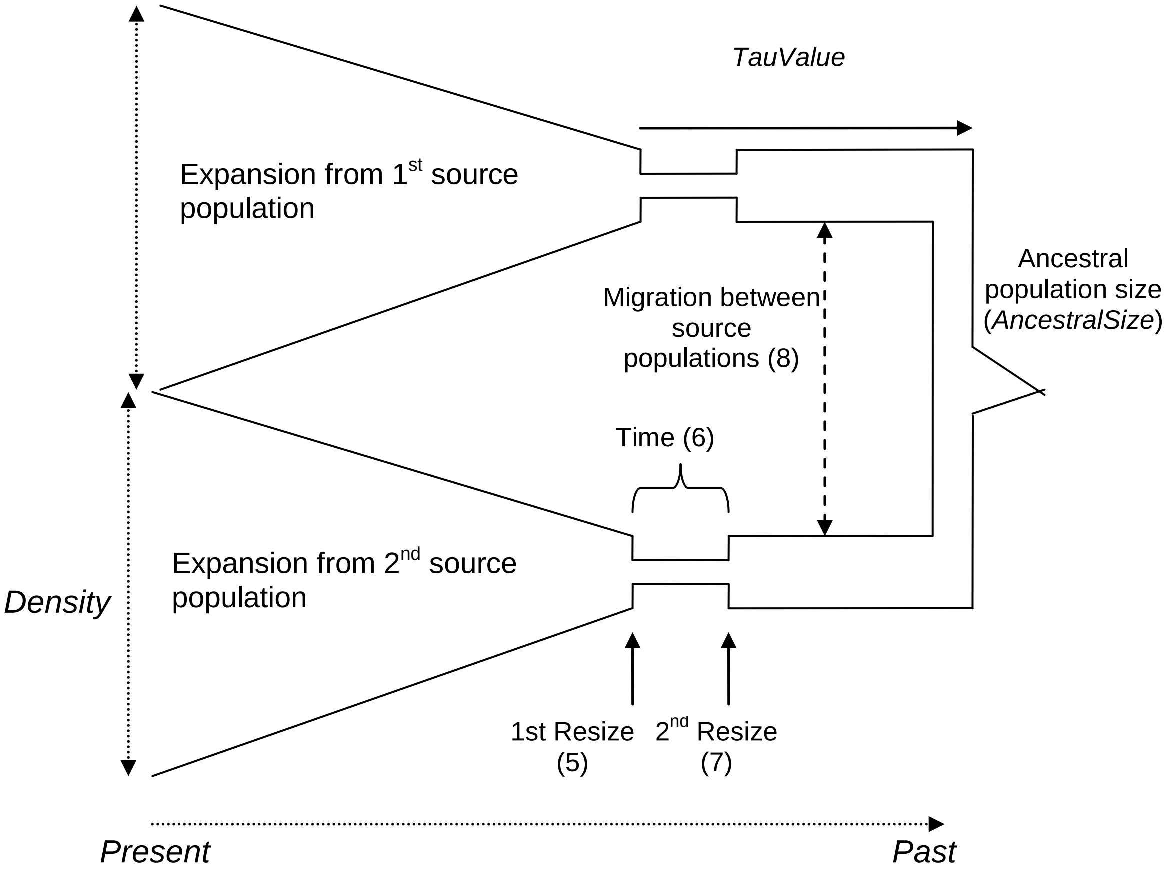

TauValue
Expansion from 1st source population
Ancestral population size (AncestralSize)
Migration between source populations (8)
Time (6)
Expansion from 2nd source population
Density
1st Resize (5)
2nd Resize (7)
Present
Past

Figure 15 Schematic representation of two parallel expansions in one layer (P1)

5.3.2 PresVegetationFile, HydroFile, RoughnessTopoFile

The environmental datasets define the "world" (size, continent contours, and geographic coordinates) in which the simulations take place. The values of these settings are the paths to the data sets that are text files in ASCII raster format. This file begins with a header (first six lines) containing information on the file, which is followed by a matrix of values in rows and columns.

The header information is as follow:

```
ncols          : number of columns
nrows          : number of rows
xllcorner      : longitude coordinate of the lower-left deme
yllcorner      : latitude coordinate of the lower-left deme
cellsize       : width of a deme (cell size), in same units than the
                 coordinates
```

```
NODATA_value  : value indicating that a deme must not be considered (e.g.
                because it is in the sea)
```

Example of an environmental dataset

```
ncols         88
nrows         91
xllcorner     -19.845388
yllcorner     -36.897187
cellsize      0.83
NODATA_value  -9999
-9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 7 7 7 ...
-9999 -9999 -9999 -9999 -9999 -9999 -9999 7 7 7 7 7 7
-9999 -9999 -9999 -9999 -9999 -9999 7 7 7 7 7 7 7 7 7
-9999 -9999 -9999 -9999 -9999 -9999 7 7 7 7 7 7 7 7 7
-9999 -9999 -9999 -9999 -9999 7 7 7 7 7 7 7 7 7 7 7
-9999 -9999 -9999 -9999 -9999 7 7 7 7 7 7 7 7 7 7 7
...
```

This format is quite standard in many GIS package, and can be exported from many of them (e.g. ArcView, ArcGIS).

The “PresVegetationFile” dataset defines vegetation categories (integer numbers) for the world (or for the first population in case of two-population setting). For each vegetation category, a carrying capacities and a friction can be defined through the Veg2K and Veg2F files (see next Section). The “HydroFile” defines rivers with any value (e.g. "1" or "0") values and non-river demes with “NoData” values. The “RoughnessTopoFile” defines relative continuous friction values, such as friction computed from topography.

Note that the RAM necessary to run SPLATCHE2 is proportional to the number of demes set through the *PresVegetationFile* data set and to the number of simulated generations. If the program crashes due to RAM limitation, either the resolution of the *PresVegetationFile* data set or the number of generations must be reduced.

5.3.3 Veg2K and Veg2F files

Carrying capacities and frictions can stay unchanged (static) during a demographic simulation, or they can vary. The way this variation is implemented is by loading other sets of carrying capacitiy and friction values at user-defined times during the simulation. If this happens, the simulation is called "dynamic". In order to set at what time(s) the changes occur, different files are needed.

The two main files, which are set in the Settings file through the settings *Veg2KFile* and *Veg2KFile,* are typically called "Dynamic_K.txt" and "Dynamic_F.txt". On the first line of each of these files appears the number of changes during a simulation. Then each line (one per change) includes the time of change (in generations), the filename of the corresponding table (see below), and an arbitrary description (spaces are allowed here). The three components of each line must be separated by a white space. For a non-dynamic simulation (0 or 1 on the first line), only the first filename is considered.

Example of "Dynamic_K.txt" file:

```
2
0 ./dataSets_africa/veg2K.txt vegetation at time 0
500 ./dataSets_africa/veg2K_500.txt doubling of vegetation at time 500
```

Each file name must target a valid "association table" that makes the link between a particular vegetation category and a carrying capacity (or friction) value.

With several tables for the carrying capacity and/or the friction values, it is possible to simulate complex changes in the environment through time.

An association table lists a series of vegetation category numbers, followed by a carrying capacity value (integer values greater than zero), and by a description. The vegetation category numbers must correspond to the numbers found in the input "vegetation" dataset (see Section 5.3.2). Note that the description of each category in the third column is mandatory (the program crashes if it is absent)!

Example of "veg2K.txt" file:

```
1      200    Tropical_rainforest
2      200    Monsoon_or_dry_forest
3      500    Tropical_woodland
4      500    Tropical_scrub
5      100    Tropical_semi_desert
6      1000   Tropical_grassland
7      50     Tropical_extreme_desert (50)
8      1000   Savanna
```

Similar files (Dynamic_F.txt) are required for setting friction values. However, this file is not mandatory. If not defined, friction values are uniform over the world. Friction values (*f*) are relative values between 0 (low friction, strong migration) and 1 (extreme friction, no migration possible). If $f \leq 0$ or $f \geq 1$, the deme is considered isolated and no migrant will reach it.

Example of "veg2F.txt" file:

```
1      0.8    Tropical rainforest
2      0.8    Monsoon or dry forest
3      0.5    Tropical woodland
4      0.5    Tropical thorn scrub and scrub woodland
5      0.2    Tropical semi-desert
```

5.3.4 MigrationRate

This is the percentage/fraction (m) of the population emigrating at each generation, and it used in the demographic models. In other words, when used in stochastic demographic models, it is the per generation probability for any gene to emigrate. For a deme population of size N the number of emigrants is then equal to $N \times m$ at each generation. Note that numbers of emigrants are always integer numbers, so the effective number of emigrants can vary slightly from the exact *N*×*m* value.

5.3.5 AllowSourcePopulationOverflow

If this parameter is set to 0, and if the initial population size (*N*) is greater than the carrying capacity (*K*) of the source population, the remaining *N-K* individuals are spread around the neighboring demes (without ever exceeding the carrying capacities of these demes) until N individuals are placed. The "overflow" of individuals allows one to start with a "patch" of source populations, rather than a single source deme. Also note that the “Resize parameter” (number 5 in section 5.3.1) should be adjusted accordingly (if not, the population size before expansion is still assumed to be equal to the carrying capacity of the source population). If this parameter is set to 1, all individuals are found in the original deme (even if N >>K), but in this case the population of the source deme can be quickly downward regulated by the logistic growth function.
Note that overflow can occur only in the first layer P1, not in the second one P2.

5.3.6 TauValue and AncestralSize

TauValue is the time before the expansion (in generations) at which the multiple source populations are pooled together into one initial population of size *AncestralSize*. These two settings are therefore only used with multiple source population simulations. Note that *TauValue* was expressed in years in previous versions of SPLATCHE. See the illustration under Section 5.3.1.

5.3.7 ArrivalCellFile

This file allows the user to specify the coordinates of the demes for which colonization times are required. An output files called “Arrival_cell_output.txt” lists the arrival time of the demes listed in the colonization file *.col.

This file follows a similar format than the initial population density file (section 5.3.1). The number of demes for which one wants information about their colonization time is given on the first line. If this number is set to 0, then nothing is done. Then, for each line the following information is given:

1. Name of the deme.

2. At which population layer the deme belongs. 0 if only one layer is simulated.

3. & 4. Latitude and longitude of the deme

Example of "colonization.col" file in case of a two-population model:

```
6
Israel        0      4      60
MiddleEast    0      12     60
Iberia        0      12     10
Turkey        1      4      60
MiddleEast    1      12     60
Iberia        1      12     10
#PopName      #Layer #Lat    # Long
```

5.3.8 FrictionChoice

Parameter allowing choosing how the friction values are computed, (0: vegetation, 1: roughness topography, 2: both). When "vegetation" or "roughness" is chosen, friction values are only computed from the corresponding input data set (see Section 5.3.2). If "both" are chosen, friction values are computed by taking, for each deme, the mean value between the friction value from the vegetation data set and the friction value from the roughness data set.

5.3.9 SampleFile

Path to a file with the extension “.sam” allowing one to specify the localization of the sampled sub-populations, as well as the number of genes sampled in each of these sub-population.

On the first line of this file, the user must specify the number (integer) of sampled sub-populations. The second line is reserved for the legends for each column. Then, each line defines a sample with 5 fields separated by “tab” or “space” character.

1. Name of the sampled sub-population.

2. Number of genes belonging to that sample.

3. Identification of the population layer to which the sampled deme belongs (0 if only one population is simulated: 1 to set it to the second population layer).

4 & **5** Geographic location of the sampled deme (latitude and longitude). As for the source deme coordinates (see Section 5.3.1), the sampled demes must fall within the extent of the world defined by the environmental files and they should not fall in "sea" (NoData) demes. If one or several sampled-subpopulation fall in the "sea", an error message will appear in the log file (and as a message box in the graphical version), but the simulation will still take place with the reminding sampled sub-populations.

Example of a genetic input file (.sam) for 6 samples in Africa:

```
6
#Name         #Size  #PopLayer     #Lat   #Long
sample1       30     0             20     20
sample2       25     0             20     0
sample3       28     0             0      20
sample4       32     0             -20    20
sample5       30     0             -30    25
sample6       40     0             5      40
```

5.3.10 GeneticFile

Path to a text file (*.par) containing the definition of markers property and recombination parameters. The format of this file is identical to the genetic marker section of SIMCOAL2 (Laval and Excoffier 2004) input files.

Here is an example of genetic file for two chromosomes (independent loci) constituted by DNA sequences of 100 base pairs, no recombination and various mutation rates:

```
2 //Number of independent chromosomes
#chromosome 1
1
DNA    100   0.0    0.001   0.8
#chromosome 2
1
DNA    100   0.0    0.0000001   0.8
```

Each line corresponds to one of the following elements:

1. Number of independent chromosome segments: The first line refers to the number of independent chromosome segments that needs to be simulated. It is possible to simulate several chromosomes with different kind and numbers of markers, different patterns of recombination rates and different mutation rates. Note however that ARLEQUIN/ARLSUMSTAT (Excoffier and Lischer 2010) is not able to analyze various kinds of markers in the same file.

2. Number of blocks: For each chromosome, the number of blocks must be defined. This number refers to the number of sets of partially linked loci (that we shall call blocks, hereafter) of the same data type, having the same mutation and recombination rates. For instance, to simulate *L* partially linked microsatellites with varying recombination rates and with different mutation rates, we need to specify *L* blocks, each one containing one microsatellite. To simulate a chromosome segment with *L* partially linked microsatellites uniformly spaced (with identical recombination rates between adjacent loci) and having identical mutation rates, one would only need to define a single block.

3. Block specificities: For every block, we need to specify 3 mandatory parameters, as follows:

3.1. Data type
3.2. Number of loci to simulate
3.3. Recombination rate immediately to the right of each locus.

In addition to these 3 required parameters, several optional parameters need to be added depending on the data type**.**

Optional parameters

	Extra parameters		
Data types	3.4	3.5	3.6
DNA	Mutation rate per locus	Transition rate (fraction of substitutions that are transitions)	
RFLP	Mutation rate per locus		
STR (microsatellite)	Mutation rate per locus. A pure stepwise mutation model is used with possible range constraints	Value of the geometric parameter for the GSM model. Ranges between 0 and 1, a value of 0 implies no insertion/deletion of more than one repeat (strict SMM model)	Range constraint (number of different alleles allowed) 0 means no range constraint
SNP	Minimum frequency for the derived allele		

In summary, the parameters which can be given for every markers are:

STR *[number of loci] [recombination rate] [mutation rate] [Geometric parameter] [range constraint]*

SNP *[number of loci] [recombination rate] [minimum frequency]*

DNA *[number of loci] [recombination rate] [mutation rate] [transition rate]*

RFLP *[number of loci] [recombination rate] [mutation rate]*

[number of loci] : Number of partially linked loci in a block
[recombination rate] : recombination rate between adjacent markers in a block (Global recombination rate of the block = [recombination rate] *([number of loci] -1))
[mutation rate] : mutation rate for every marker in a block. (Global mutation rate of the block = [mutation rate] *[number of loci])
[Geometric parameter]: parameter *p* for Generalized Stepwise Mutation *GSM(model . The number of repeats inserted (deleted) in one mutation event is treated as a random variable, and is randomly drawn from a Geometric distribution with parameter *p*. It may vary between 0 and 1 (0 means that the number of repeats

inserted or deleted for **every** mutation event is 1 while 0.5 means that the **expected** number of repeats is 2. See http://cmpg.unibe.ch/software/simcoal2/#GSM for more details.
[range constraint] : 0 means no constraints, higher numbers give the maximum possible number of repeats
[minimum frequency] : minimum frequency of a SNP in the whole set of population (ascertainment bias)
[transition rate] : fraction of substitutions that are transitions

Additional examples of genetic input files can be found in the web site of the program SIMCOAL2: ***http://cmpg.unibe.ch/software/simcoal2***

5.3.11 Coalescent distribution

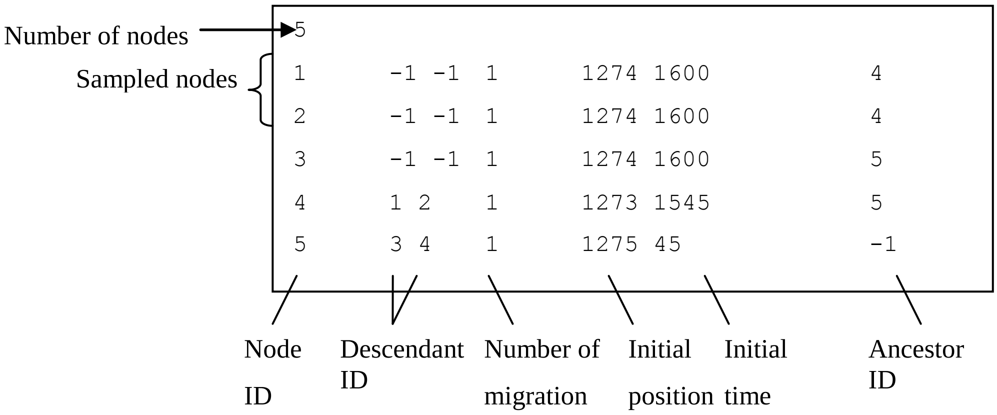
In settings the checkbox "Generate bitmap" on the graphical interface, it is possible to generate (only for the last chromosome simulated) bitmaps representing the location of genes either over densities of layer1 or layer2 (choice made in the "General Output" panel) or alternatively over occupation. Bitmaps can be generated every *n* generations in color or grayscale. Those file names start with *".genetic"* and end with *".bmp"*.

5.3.12 Coalescent tree files

In settings the checkbox "coalescent trees" on the graphical interface, it is possible to generate for each simulation a bitmap representing the branches of the coalescence tree for the last chromosome simulated, laid out spatially. These files are terminated with *"*_CoalTree_*.bmp"*. This feature is only available with one population layer and without recombination (the file is not generated when recombination is used). If more than one chromosome is simulated, the file is generated only for the last chromosome without recombination.

5.3.13 GenealogiesFile

This settings controls whether a *.tri file(s) is generated to store the geographic position of the coalescent events having occurred in the coalescent process, as well as the migration of the nodes between demes. These files (one per chromosome) are only generated when no recombination is used. With one layer, the *tri files looks as follow:

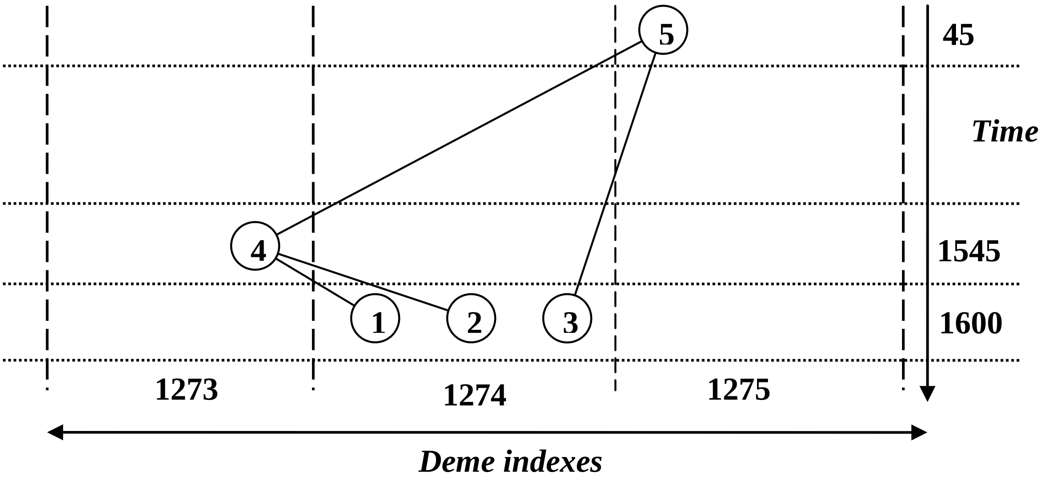


```
5
1      -1 -1  1      1274 1600           4
2      -1 -1  1      1274 1600           4
3      -1 -1  1      1274 1600           5
4      1 2    1      1273 1545           5
5      3 4    1      1275 45             -1
```

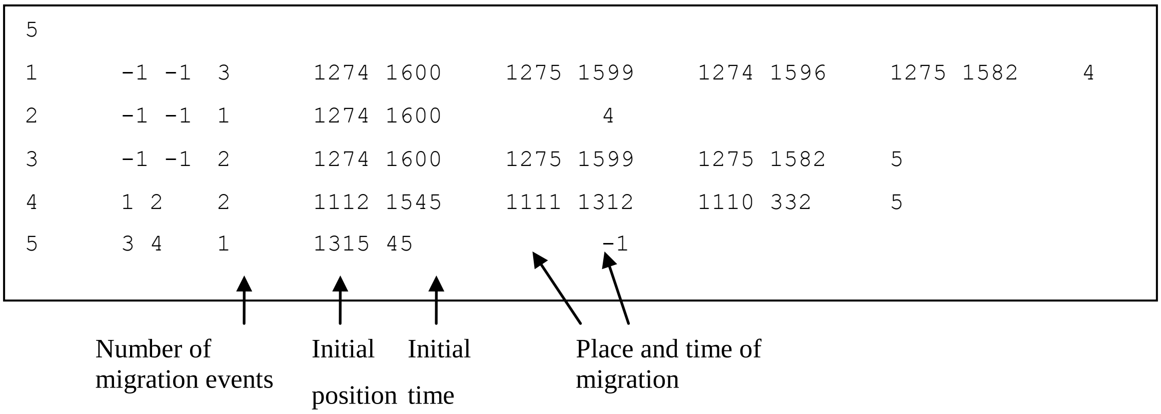
The first column of the first line corresponds to the number *n* of nodes (5 in this case, the first three nodes are sampled genes). For each node, with IDs between 1 and *n*, corresponds a line where the numbers are:

1. Node ID
2. Descendent nodes ID (-1 if the node is sampled)
3. Number of migration events for this node. This number is always equal to 1, and thus uninformative except when the parameters *DoublePopulationMode* and *PropFile* are both set to 1 (see below).
4. Index of the cell in which the node appears.
5. Time (in generations) at which the node appears. In the case of our example, the simulation ends after 1600 generations, which is the current sampling time.
6. Ancestor node ID (-1 for the MRCA)

For the *.tri file example mentioned above, see below a schematic representation of the corresponding coalescent tree displayed over space and time.



In the particular case of two populations (*DoublePopulationMode*=1) and if the setting *PropFile* is set to 1, then every node migrations are stored into the file as follows:



```
5
1      -1 -1  3      1274 1600     1275 1599     1274 1596     1275 1582     4
2      -1 -1  1      1274 1600            4
3      -1 -1  2      1274 1600     1275 1599     1275 1582     5
4      1 2    2      1112 1545     1111 1312     1110 332      5
5      3 4    1      1315 45              -1
```

This example shows that node 1 migrated 3 times, from deme 1274 to deme 1275 at time 1599, from 1275 to deme 1274 at time 1596 and finally from deme 1274 to deme 1275 at time 1582, etc…

5.3.14 ImmigDistrFile

This setting controls whether a file is created with the number of immigrants in sampled demes for every generations. One file ”*.immigrants_DemeIndex.nm” is created per sampled deme.

The first column corresponds to the generations and the second one to the number of individual that arrived in the current deme generation after generation.

```
Generation Immigrants
1      0
2      0
3      0
4      2
5      3
6      5
7      8
8      8
9      8
```

5.3.15 Admixture rate from P1 to P2 and Admixture rate from P2 to P1

This is the rate of gene flow from demes belonging to the first population layer (P1) into demes belonging to the second population layer (P2) or the reverse. It represents admixture or hybridization and could only occur between demes belonging to an identical geographical cell (same location in both arrays of demes). This parameter corresponds to γ_{ij} in the equations in section 4.5. This admixture rate can be set between 0 (no admixture) and 1 (full interbreeding). If the value of *Admixture rate* *from P1 to P2* is different from the value of *Admixture rate from P2 to P1* then the admixture between both layers is asymmetrical.

5.3.16 PropFile (*.prop)

When simulating two populations (*DoublePopulationMode*=1), it is possible to generate a file that gives the proportion of genes sampled in layer two (P2) that comes originally from the source deme of that layer. In other words, this is the proportion of genes in the invasive population P2 which have not been introgressed by local P1 genes. See (Currat, Ruedi et al. 2008) for more details.

One file is generated for each simulation. In each file, there is one line per chromosome in which the proportions of genes sampled in layer two (P2) that comes originally from the source deme of that layer, are given for all sampled genes taken together, but also for each sample separately.

```
Proportion of lineages sampled in P2 which derived from the P2 source
population

#Chrom Total_prop   1      2      3      4      5

Chrom_1      0.415         0.5    0.3    0.45   0.65   0.4

Chrom_2 Not available with recombination!
```

5.4 Other generated outputs

5.4.1 Log files

When running SPLATCHE2, information and potential errors are written in a log file called "splatche2_win.log" for the graphical Windows version, and "splatche2.log" for Windows and Linux console versions.

5.4.2 ARLEQUIN output file

The genetic data generated by one simulation are directly output in an ARLEQUIN project file, with the extension "**.arp*". This file format allows one to analyse the data with the ARLEQUIN program in order to obtain different statistics, see ARLEQUIN manual (Excoffier and Lischer 2010) and website (http://popgen.unibe.ch/software/arlequin35) for more details. If more than one simulation is performed using one demographic simulation (which is usually the case) then an ARLEQUIN batch file (with extension "**.arb*") is also generated, listing all simulated files, and allowing one to compute statistics on the whole set of simulated files. Note also that the ARLEQUIN software has a file conversion utility for exporting input data files into several other format like BIOSYS, PHYLIP, or GENEPOP, so that files produced by SPLATCHE2 could be also analyzed by these software after file conversion. Alternatively, PGDSpider (http://www.cmpg.unibe.ch/software/PGDSpider/), a program developed by the CMPG lab, can also be used to convert population genetic data between many formats.

Note that different types of markers can be simulated at the same time in SPLATCHE2 (see how to do it in Section 5.3.10). These simulated markers will be output in the ARLEQUIN files, but the current version of ARLEQUIN (ver. 3.5) is not able to read and further analyze different set of markers. The user of such an output file would therefore need to develop his/her own routines to treat separately the different sets of markers.

The output ARLEQUIN files will be located in a folder called "GeneticsOutput" located in the folder holding the *.sam and *par files.

5.4.3 Coalescence distribution

With the graphical version, a bitmap representing the spatial distribution of the coalescent events over all simulations (and chromosomes) is automatically created with the "*_TotNumCoal.bmp" suffix. This bitmap can also be visualized by means of the button "Draw Coalescence" on the graphical interface. The bitmap is only created in absence of recombinations.

By checking the "coalescent spatial distribution" checkbox on the graphical interface, similar bitmaps of the spatial distribution of coalescent events are generated for every simulation (with the "*_NumCoal_simi_chromj.bmp" termination).

By checking the "coalescent temporal distribution" checkbox on the graphical interface, times for each coalescent event and each simulation are listed on a file with "*.coal" extension. Those times are indicated in generation units, with the largest number corresponding to the end time of the simulation.

5.4.4 MRCA files

With the graphical version, SPLATCHE2 also gives information on the location and the timing of the Most Recent Common Ancestors (MRCAs) of the sampled genes,. A bitmap file with the termination "*_TotMRCADensity.bmp" is automatically generated and shows the spatial distribution of MRCA for all simulations. These maps can also be visualized by checking the button "Draw MRCA" on the graphical interface. These maps are only created when no recombination is used.

5.4.5 Arrival cell files

A file called "Arrival_cell_output.txt" is generated, listing the arrival time (in generation) in the cells specified in the "*.col" input file.

5.4.6 Nexus

Two other types of file produced by SPLATCHE are compatible with the NEXUS tree file format: for each simulation, a file with "*.paup" extension could be generated. This file lists all the simulated genes together with their true genealogical structure. This file can be analyzed by many programs (e.g. with David Swofford's PAUP* software http://paup.csit.fsu.edu). A batch file, with extension "*.bat" is also generated.

6 References

Currat, M. and L. Excoffier (2004). "Modern humans did not admix with Neanderthals during their range expansion into Europe." PLoS Biol **2**(12): 2264-2274.

Currat, M. and L. Excoffier (2005). "The effect of the Neolithic expansion on European molecular diversity." Proc Biol Sci **272**(1564): 679-88.

Currat, M., L. Excoffier, 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**(5784): 172; author reply 172.

Currat, M., N. Ray, et al. (2004). "SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity." Molecular Ecology Notes **4**(1): 139-142.

Currat, M., M. Ruedi, et al. (2008). "The hidden side of invasions: massive introgression by local genes." Evolution **62**(8): 1908-20.

Excoffier, L. and H. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." Molecular Ecology Resources **10**: 564-567.

Excoffier, L. and H. E. L. Lischer (2010). "Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows." Molecular Ecology Resources. **in press**.

Foll, M. and O. Gaggiotti (2006). "Identifying the Environmental Factors That Determine the Genetic Structure of Populations." Genetics **174**(2): 875-891.

Francois, O., M. Currat, et al. (2010). "Principal component analysis under population genetic models of range expansion and admixture." Mol Biol Evol **27**(6): 1257-68.

Hamilton, G., M. Currat, et al. (2005). "Bayesian estimation of recent migration rates after a spatial expansion." Genetics **170**(1): 409-17.

Hewitt, G. M. (2004). "Genetic consequences of climatic oscillations in the Quaternary." Philos Trans R Soc Lond B Biol Sci **359**(1442): 183-95; discussion 195.

Hudson, R. (1990). Gene genealogies and the coalescent process. Oxford, Oxford University Press.

Kimura, M. and W. H. Weiss (1964). "The stepping stone model of genetic structure and the decrease of genetic correlation with distance." Genetics **49**: 561-576.

Klopfstein, S., M. Currat, et al. (2006). "The fate of mutations surfing on the wave of a range expansion." Mol Biol Evol **23**(3): 482-90.

Laval, G. and L. Excoffier (2004). "SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history." Bioinformatics **20**(15): 1485-2487.

Lotka, A. J. (1932). "The growth of mixed populations: two species competing for a common food supply." Journal of the Washington academy of Sciences **22**: 461-469.

Quemere, E., B. Crouau-Roy, et al. (2010). "Landscape genetics of an endangered lemur (Propithecus tattersalli) within its entire fragmented range." Mol Ecol **19**(8): 1606-21.

Ray, N. (2003). Modélisation de la démographie des populations humaines préhistoriques à l'aide de données environnementales et génétiques.

Départment d'Anthropologie. Genève, Université de Genève, available at http://www.unige.ch/cyberdocuments/theses2003/RayN/meta.html**:** 331.

Ray, N., M. Currat, et al. (2005). "Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations." Genome Research **15**(8): 1161-7.

Ray, N., M. Currat, et al. (2003). "Intra-deme molecular diversity in spatially expanding populations." Molecular Biology and Evolution **20**(1): 76-86.

Taberlet, P., L. Fumagalli, et al. (1998). "Comparative phylogeography and postglacial colonization routes in Europe." Molecular Ecology **7**(4): 453-464.

Volterra, V. (1926). Variations and fluctuations of the numbers of individuals in animal species living together (Reprinted in 1931). Animal Ecology. R. N. Chapman. New York, Mc Graw Hill**:** 409-448.

Wegmann, D., M. Currat, et al. (2006). "Molecular diversity after a range expansion in heterogeneous environments." Genetics **174**: 2009-2020.