

# Documentation v0.7

<b>I.</b>	<b>HOW TO USE AND RUN THE PROGRAM .....</b>	<b>2</b>
1.	Installation .....	2
2.	Running simulation examples.....	2
3.	Building simulations .....	2
3.1.	Working directory .....	3
3.2.	Using input files .....	4
3.3.	Defining the landscape in the GUI .....	4
3.4.	Running simulation examples .....	5
3.5.	Defining the individuals characteristics .....	5
3.6.	Defining the landscape changes over time .....	5
3.7.	Defining output files.....	5
3.8.	Running the simulation .....	6
4.	Running SimAdapt from R software .....	6
<b>II.</b>	<b>ODD PROTOCOL.....</b>	<b>8</b>
1.	Purpose .....	8
2.	Entities, state variables, and scales .....	8
2.1.	Individuals .....	8
2.2.	Grid cells .....	8
3.	Process overview and scheduling .....	10
4.	Design concepts .....	10
4.1.	Basic principles .....	10
4.2.	Emergence .....	10
4.3.	Adaptation .....	10
4.4.	Objectives.....	10
4.5.	Learning .....	11
4.6.	Prediction .....	11
4.7.	Sensing .....	11
4.8.	Interaction.....	11
4.9.	Stochasticity .....	11
4.10.	Collectives.....	11
4.11.	Observation .....	11
5.	Initialization.....	11
6.	Input data .....	14
7.	Submodels .....	14
7.1.	Individuals related submodels .....	14
7.2.	Initialization related submodels.....	16
7.3.	Landscape related submodels .....	16
7.4.	Output.....	17
<b>III.</b>	<b>VALIDATION.....</b>	<b>18</b>
1.	Validation by comparing SimAdapt simulation results including selection with population genetic theoretical expectations .....	18
2.	Validation by comparing SimAdapt simulation results without selection with population genetic theoretical expectations in a single panmictic population .....	20
3.	Validation by comparing SimAdapt simulation results without selection with population genetic theoretical expectations in an island model .....	21
<b>IV.</b>	<b>EXAMPLE FILE “HABITAT_TYPE.TXT” .....</b>	<b>23</b>
	<b>REFERENCES .....</b>	<b>25</b>

## I. HOW TO USE AND RUN THE PROGRAM

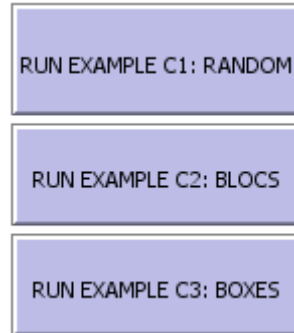
### 1. Installation

Install NetLogo (Wilensky 1999), freely available at <http://ccl.northwestern.edu/netlogo/>. We used NetLogo 5.0.3 to develop this simulation model, which runs under any operating systems with a Java virtual machine (Linux, MS Windows, Mac OS X). 1024 Mo of RAM at least is needed for the model to run properly.

Open NetLogo, and then open the simulation model using File / Open... / program.txt. Alternatively you can open the file “*simadapt\_v20\_07*” using NetLogo. If the model is properly loaded, you should see “*simadapt\_v20\_07 – NetLogo {path to the model directory}*” at the top left of the current window.

### 2. Running simulation examples

The three simulation examples presented in the main text are available using the corresponding buttons “*RUN EXAMPLE C1*”, “*RUN EXAMPLE C2*” and “*RUN EXAMPLE C3*”.



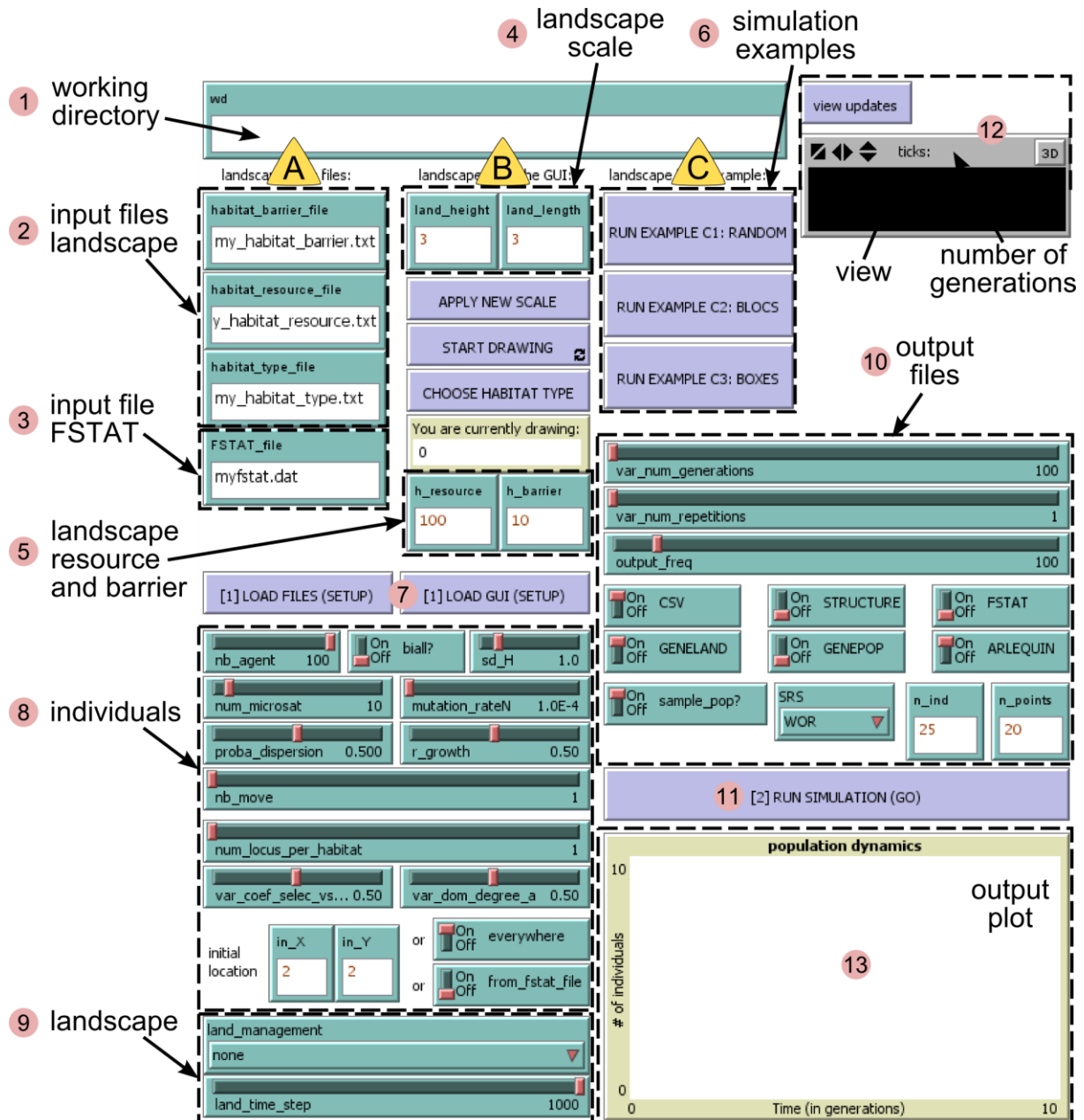
**Fig. S1:** buttons to run the simulation examples in the GUI

The simulations are performed once (no repetition), and take less than 40 seconds each using Windows Vista 64bits, CPU 3.06GHz, RAM 4Go (“view updates” option disabled by default to optimize computational speed). The output files are created in the directory where the program is (unless a different working directory has been specified in the input `wcd`). The simulation can be repeated using the BehaviorSpace (Tools/BehaviorSpace). If your computer has multiple processor cores, then model runs could happen in parallel (number of cores defined by the user). The program can also run headless (without graphical user interface) in a distant server or a cluster of machines. More information and NetLogo documentation on BehaviorSpace can be found at the following URL:

<http://ccl.northwestern.edu/netlogo/docs/behaviorspace.html>

### 3. Building simulations

In this section, we described all the features of the program and parameters available in the graphical user interface (or headless using *BehavriorSpace*).



**Fig. S2: Representation of the graphical user interface**

### 3.1. Working directory

As simulations can produce a large number of files, the first step is to define the working directory `wd` (Fig. S2, box 1) where the output files will be generated. If nothing is mentioned, the current directory is used (directory of the NetLogo model file).




Using MS Windows, an example could be:

`C:\path\to\my\directory\`


Using Linux or Mac OS X:


`/home/path/to/my/directory/`


Please note that unlike our example, a working directory like C:\path\to\my\directory will produce output files in the C:\path\to\my\directory with output files beginning with directory.

Once the working directory is defined, you can either: (Fig. S2, box ) generate a landscape from input files, (Fig. S2, box ) draw a landscape from the GUI or (Fig. S2, box ) run the simulation examples.



### 3.2. Using input files (Fig. S2, box )

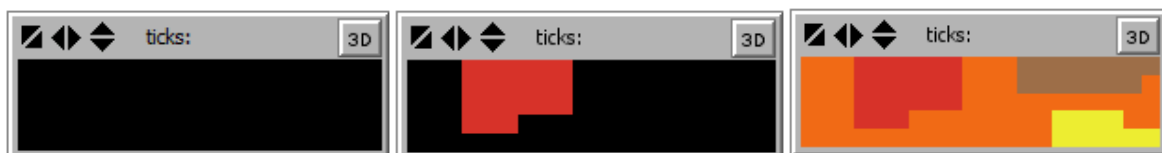
Landscape input files (Fig. S2, box ) define the three layers of the landscape : the matrix of resistance for the individual dispersion (*habitat\_barrier\_file*), the carrying capacity for individuals (*habitat\_resource\_file*), and the different habitat types (*habitat\_type\_file*). Those files are composed of three columns: the abscissa (x from 0 to X, *i.e.* X+1 grid cells), the ordinate (y from 0 to Y, *i.e.* Y+1 grid cells) and the value of the corresponding layer. An example is provided in the third part of this supplementary material (see III. Example file *habitat\_type.txt*). Files must be located in the same working directory defined earlier, and their name should include the extension *.txt*.

Once your files are created, you can load them into the model using the “[I] LOAD FROM FILES” button in the GUI (Fig. S2, box ). A message will sum up the information loaded and check the correctness of the input files.



FSTAT input files (Fig. S2, box ) allows you to load neutral markers information from an FSTAT file.

### 3.3. Defining the landscape in the GUI (Fig. S2, box )

In case you want to define your own landscape files, the first thing to do is to define the scale of your landscape (Fig. S2, box ). The scale is defined by the number of cells in abscissa (*land\_lenght*), and in ordinate (*land\_height*). A reasonable landscape for a simulation running on a laptop should be around 100 cells (*i.e.* 5\*20 or 10\*10). The “APPLY NEW SCALE” button updates the landscape scale (Fig. S2, box ). At this point you should verify that the “view updates” checkbox is checked (see above), to visualize the landscape. You can then use the “START DRAWING” button.



**Fig. S3:** Steps during the drawing of a user-defined landscape.


Then choose (Fig. S2, box ) the carrying capacity (*h\_resource*) and resistance (*h\_barrier*) associated with your first habitat type, and select your habitat type using “CHOOSE HABITAT TYPE” button. Just click in the view (Fig. S2, box ) to affect your settings to a given coordinate, and repeat the operation until all locations are set (a minimum

of two habitat types is required, see Fig. 2). Finally, use the “[1] LOAD FROM GUI” button to validate your landscape. A similar text should be displayed in the Command Center:

```
Your landscape has been correctly initialized
Average landscape barrier: 10
Average landscape resource: 100
Number of habitat types: 4
```

A copy of your three input files is now located in your working directory, which can be used later as input files (*my\_habitat\_type.txt*, *my\_habitat\_barrier.txt*, *my\_habitat\_resource.txt*). You should now uncheck the “view updates” checkbox to optimize computation time.

### 3.4. Running simulation examples (Fig. S2, box )

The simulation examples (Fig. S2, box ) can be replicated using the corresponding buttons “RUN EXAMPLE C1”, “RUN EXAMPLE C2” and “RUN EXAMPLE C3”. Only one repetition is performed, which should take less than a minute in almost all computers.

### 3.5. Defining the individuals characteristics (Fig. S2, box )

*nb\_agent*: number of individuals at initialization.  
*in\_X*: abscissa for the initial location of individuals.  
*in\_Y*: ordinate for the initial location of individuals.  
*everywhere*: switch to ON to create *nb\_agent* in each location of the landscape.  
*from\_fstat\_file*: switch to ON to generate individuals from an FSTAT file instead of using *nb\_agent*.  
*num\_microsat*: number of microsatellites loci for each individual.  
*sd\_H*: standard deviation of the normal distribution defining the heterozygosity rate and the number of different alleles (see input section of the ODD protocol below).  
*biall?*: switch to choose biallelic neutral loci instead of *sd\_H* (alleles “1” and “0”).  
*mutation\_rateN*: probability of a stepwise mutation event.  
*proba\_dispersion*: probability for an individual to disperse to another location.  
*nb\_move*: maximum distance of dispersion for an individual (in number of cells).  
*r\_growth*: growth parameter of the logistic growth submodel.  
*num\_locus\_per\_habitat*: number of loci under selection for each habitat type.  
*var\_coef\_selec\_vs...*: selection coefficient against the disfavored genotype (see Hartl and Clark, 2007).  
*var\_dom\_degree\_a*: degree of dominance (see Hartl and Clark, 2007).

### 3.6. Defining the landscape changes over time (Fig. S2, box )

*land\_management*: type of land management.  
*land\_time\_step*: frequency of landscape management (every x generations).

### 3.7. Defining output files (Fig. S2, box )

*var\_num\_generations*: number of generations for the simulation.  
*var\_num\_repetitions*: number of repetitions for the simulation.  
*output\_freq*: frequency of output files creation (every x generations).

CSV: switch to create an output file of this format (never sampled).  
GENELAND: switch to create an output file of this format.  
STRUCTURE: switch to create an output file of this format.  
GENEPOP: switch to create an output file of this format.  
FSTAT: switch to create an output file of this format.  
ARLEQUIN: switch to create an output file of this format.  
sample\_pop?: switch to sampled populations (random sampling among individuals).  
SRS: type of sampling method (WR: with replacement and WOR: without replacement)  
n\_point: number of localizations per habitat type for the sampling.  
n\_ind: number of individuals per localization for the sampling.

### 3.8. Running the simulation

Once the landscape and the individuals are set, the simulation can run using the “[2] RUN SIMULATION” button (Fig. S2, box 11). The population dynamics is shown in plot (Fig. S2, box 13).

## 4. Running SimAdapt from R software

NetLogo users is a fast growing user community and the free software is among leading platforms for the implementation of individual-based and agent-based models (see figure 2 in Le Page *et al.* 2012). Support for analysis of simulations is limited, but the statistics software R can provide such support. Thiele *et al.* 2012 developed an R package that link R and NetLogo, allowing any NetLogo program to be controlled and run from R. Bellow is an R script file which exemplified the use of SimAdapt from R software. RNetLogo package is available on CRAN (<http://cran.r-project/web/packages/RNetLogo>), along with a user manual and tutorial.

**Warning:** SimAdapt parameterization from R does not include limited range of values provided in NetLogo GUI. Of particular importance is the parameterization of  $s$  and  $h$  coefficients of selection:  $s$  should be equal or inferior than 1 and  $sh$  equal or inferior to 1.

```
#!/usr/bin/env Rscript

#script to run SimAdapt model from R: a simulation example

# loading packages
#####
#install.packages("RNetLogo", dependencies = TRUE) #linux users may need to do "R CMD javareconf" with root privileges in their terminal
before installing RNetLogo (for the rJava package)

library(RNetLogo) #package RNetLogo by Thiele et al. (2012) EMCSR (jthiele@gwdg.de)

nl.path          <- "/directory/of/netlogo/program/"
model.path       <- "/directory/of/model/simadapt_v20_07.nlogo"

# RNetLogo
#####
instance<-"instance_1"
NLStart(nl.path, gui=FALSE, obj.name=instance, nl.version=5, is3d=FALSE)
      NLLoadModel(model.path, nl.obj=get(instance))

      NLCommand("setup", nl.obj=get(instance))
      NLCommand("set land_height 2", nl.obj=get(instance))
      NLCommand("set land_length 2", nl.obj=get(instance))
      NLCommand("resize-world 0 (land_length - 1) 0 (land_height - 1)", nl.obj=get(instance))
      NLCommand("set everywhere TRUE", nl.obj=get(instance))
      NLCommand("set land_management \"none\"", nl.obj=get(instance))

      NLCommand("set var_num_generations 10", nl.obj=get(instance))
      NLCommand("set output_freq 5", nl.obj=get(instance))
      NLCommand("set ARLEQUIN TRUE", nl.obj=get(instance))

      NLCommand("ask patches [set habitat_barrier 10]", nl.obj=get(instance))
      NLCommand("ask patches [set habitat_resource 50]", nl.obj=get(instance))
      NLCommand("ask patches [set habitat_type 1]", nl.obj=get(instance))
      NLCommand("export_GUI", nl.obj=get(instance))
      NLCommand("load_input_files", nl.obj=get(instance))

      NLCommand("setup_create_turtles", nl.obj=get(instance))
      NLCommand("go_init", nl.obj=get(instance))
NLQuit(nl.obj=get(instance))

#creating an instance for NetLogo
#starting NetLogo headless (gui=FALSE)
#loading SimAdapt

#clear all from previous runs
#set landscape height
#set landscape width
#apply new scale
#set individuals everywhere
#set landscape management to "none"

#set simulation time to 10 generations
#set output frequency to 5
#set output file ARLEQUIN true

#load user-defined habitat_barrier
#load user-defined habitat_resource
#load user-defined habitat_type
#save landscape into external files
#load landscape from external files

#create individuals
#RUN SIMULATION
#quit NetLogo
```

## II. ODD PROTOCOL

The model description follows the ODD (Overview, Design concepts, Details) protocol for describing individual- and agent-based models (Grimm *et al.* 2006, 2010). It is destined to readers looking for a full description and verification of a particular process, but most of all for users willing to modify or extend the code to their own study cases.

### 1. Purpose

Worldwide, populations evolve on various habitats, in which the level of modification or alteration by human activities is heterogeneous. The aim of this model is to study the impact of human activities, through land use, on the structure of populations in order to establish scenarios including changes likely to append in land use or at a broader context, global changes.

### 2. Entities, state variables, and scales

#### 2.1. Individuals

Each individual in the model represents an entity, characterized by state variables updated at each time step (see Table S1) stored in an object `turtle` located in a given grid cell (landscape). Each time step in the model corresponds to one individual generation (non-overlapping) and simulations could run for any number of time steps.

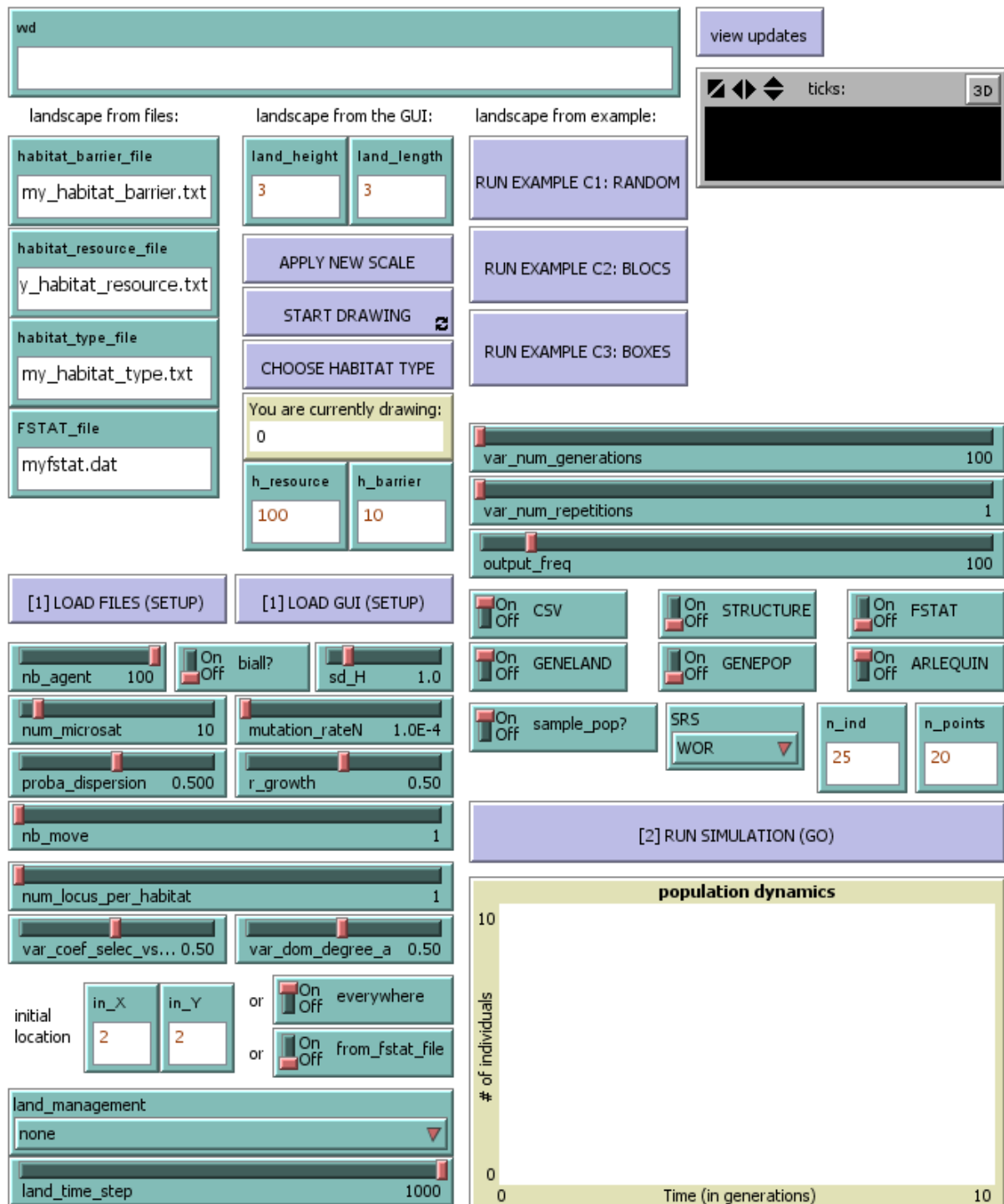
**Table S1.** State variables for individuals.

<i>State variable names in the code</i>	<i>Definition</i>	<i>Type</i>
<code>self</code>	Number to identify each individual	String
<code>cap_move</code>	Default value for individuals dispersion capabilities	Integer
<code>markerN</code>	List of neutral markers (microsatellites or SNPs)	List (Integer)
<code>markerS</code>	List of alleles for each locus under selection	List (Integer)
<code>generation</code>	Variable to separate generations (“old” or “new”)	String
<code>wi</code>	Fitness	Float
<code>mymum</code>	Genealogy	String
<code>mydad</code>	Genealogy	String
<code>sample</code>	Variable used in sampling (with or without replacement)	Boolean

#### 2.2. Grid cells

The grid represents the landscape in which the individuals evolve. Each cell is characterized by its state, descriptor of environmental conditions which drive the behavior and dynamics of individuals, composed of a barrier variable representing the cost to move to a given cell (`habitat_barrier`), a resource representing the suitability of a given patch (`habitat_resource`), i.e. the carrying capacity and a habitat type representing the characteristics regarding adaptation (`habitat_type`). The cells changes over time according to management rules. The landscape can be defined in the Graphical User Interface (GUI), see Fig. S4.





**Fig. S4:** Graphical User Interface of the simulation model using NetLogo 5.0.2 (Wilensky, 1999)

### 2.3. Collectives

Group of individuals located in the same habitat type (or in the same location), are considered as populations for further analysis.

### **3. Process overview and scheduling**

The model processes include the dispersion and reproduction of individuals in this given order for each given individual taken in a random order. Time is modeled as discrete steps. At the end of one time step, all individuals die, and the next time step begins with their offspring (*i.e.* non-overlapping generations). The information regarding the previous generation after reproduction is stored in an output file only if requested. All individual state variables are updated asynchronously as they are taken in a random order.

### **4. Design concepts**

#### **4.1. Basic principles**

The model design is based on previous work on the fields of landscape genetics where significant advances have been made on the last decade (Manel *et al.* 2003 ; Manel and Segelbacher 2009). This simulation model is generic enough to be adapted for all living forms and is based on previous works by Gavrillets *et al.* 2007 and Gravilets and Vose, 2005 for their study on *Cichlidae* ; Duenez-Guzman *et al.* 2009 for their work on *Heliconius* ; Bruggeman *et al.* 2010 and Bruggeman *et al.* 2009 for their work on *Picoides borealis* ; Philips *et al.* 2004 for their work on trees ; Lawson et Jensen, 2005 ; Saledin et Littlejohn, 2003 ; Landguth *et al.* 2010a ; Landguth *et al.* 2010b ; Landguth *et al.* 2010c and Jaquiéry *et al.* 2011 for their generic studies. Our work differ from those for its generic approach with genetic, ecological and landscape submodels with a high level of abstraction allowing to attempt the construction of scenarios rather than explaining a given situation (from real world to theoretical situations). The landscape submodel can run asynchronously with others submodels to reproduce time lags or independent landscape management. Our work differs from other forward-time simulation program such as simuPop (Peng and Kimmel 2005), Nemo (Guillaume and Rougemont 2006) or quantiNEMO (Neuenschwander *et al.* 2006) for its focus on landscape genetics with a highly flexible and extendable landscape submodel linked to the population genetics approach.

#### **4.2. Emergence**

Almost all results of the model emerged from the behavior of the individuals depending on their adaptive traits in the given changing landscape. However initial characteristics of individuals are imposed and hence dependent on what type of individual is simulated. Consequently, they are rather ‘built in’ than emergent results.

#### **4.3. Adaptation**

Adaptive traits were considered for all individuals through reproduction. Individuals with a selective advantage (*e.g.* a particular genotype) in a specific habitat are expressing a better fitness than other individuals. Applying Mendel inheritance laws and ecology concept of carrying capacity, in the next generation the proportion of individuals with high fitness will thereby be increased.

#### **4.4. Objectives**

Individuals do not make decisions by ranking alternatives and fitness is a consequence of random move and genetic background (genotype).

#### **4.5. Learning**

Fitness changes reflect adaptative process to a given habitat over time (*i.e.* no learning). The adaptative trait is heritable and subject to mutations.

#### **4.6. Prediction**

The dispersion of individuals can be considered as a tacit prediction where the dispersion to another cell will give to the individuals a better probability of survival.

#### **4.7. Sensing**

Individuals perceive their peers in a given cell and the location of potential destinations in their neighborhood. The mechanisms by which individuals obtain information are modeled explicitly.

#### **4.8. Interaction**

Interactions between individuals are direct for reproduction and indirect for mediating resources (carrying capacity).

#### **4.9. Stochasticity**

The mating of individuals in a given cell are modeled randomly to reproduce variability. The dispersion of individuals is random among possible destinations because we assumed that individuals do not have a perception of habitats located in other cells.

#### **4.10. Collectives**

Individuals located in the same grid cell (patch), belong to an aggregate named population. Population genetics analyses can consequently be made at the grid cell level. Alternatively, populations can be considered as the aggregation of all individuals in all cells belonging to the same habitat type (*e.g.* a landscape of 3 different habitat types corresponds to 3 populations). User-defined populations can also be defined using external software such as GENELAND or STRUCTURE.

#### **4.11. Observation**

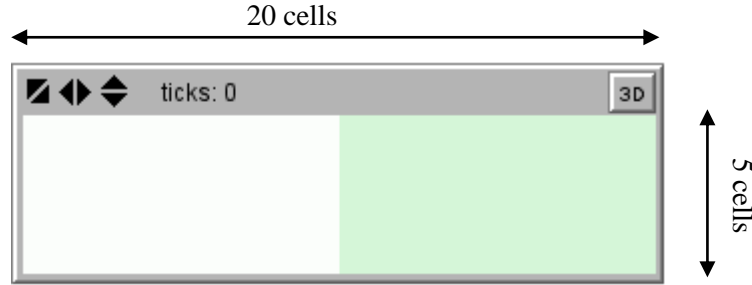
All information regarding individuals can be collected for analyzing the model after each reproduction process (*i.e.* at each time step). This information can also be sampled and used to imitate what can be observed in an empirical study (see the “Virtual Ecologist” approach by Zurell *et al.* 2010, as mentioned by Grimm *et al.* 2010).

### **5. Initialization**

The initial state of the model (*i.e.* at time  $t = 0$ ) is composed of a user-defined number of individuals located in the landscape. Alleles at both the neutral loci and loci under selection are initialized randomly according to the setting and consequently vary among simulations. In this simulation model, consequences of initial state are studied so that user-defined initialization is of importance in order the results to be accurately replicated. Default values for the simulation examples are provided in Table S2. The simulation model can use external files for the integration of the landscape layers (GIS). These files concern the landscape habitat (“*habitat\_type.txt*”), the landscape resource (“*habitat\_resource.txt*”) and the landscape barrier (“*habitat\_barrier.txt*”). Fig. S5 represents an example of input file at initialization for habitat types. If no files are provided, the model will run with default values of 1, 100 and 10,

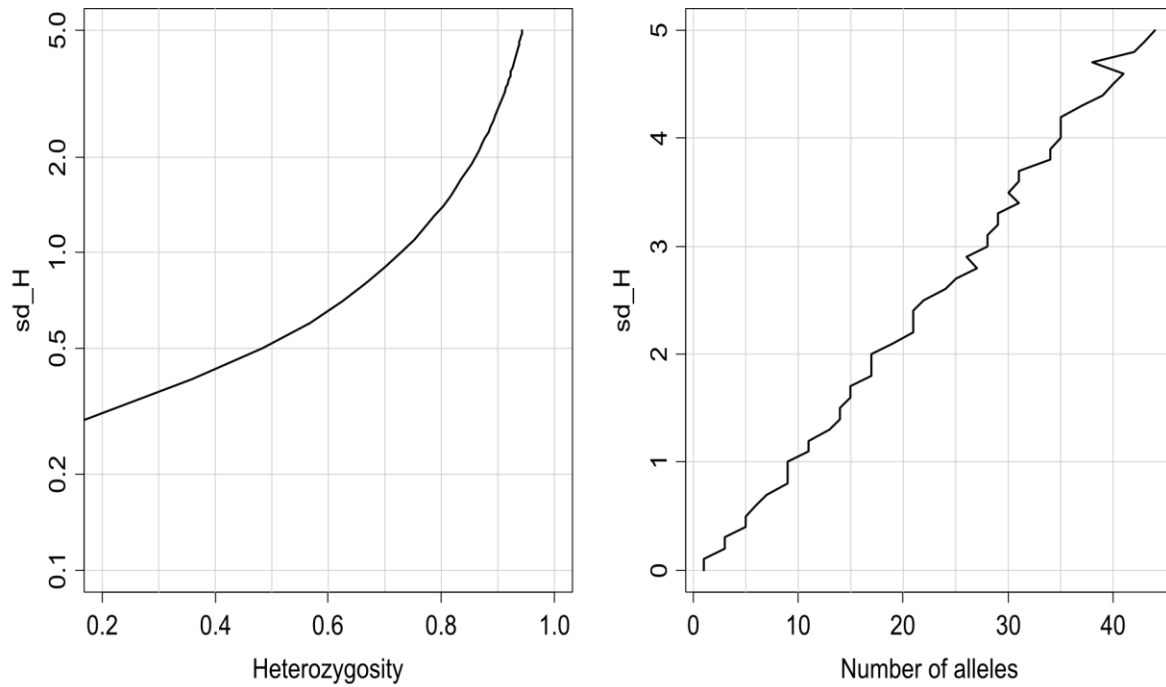
respectively. Each file is organized with coordinates in the first two columns (abscissa, ordinate) and landscape characteristic in the third as following (see also section III):

```
0    0    1
1    0    4
2    0    1
...
10   10   3
```



**Fig. S5:** Graphical representation of an example of habitat types at initialization

Alleles are chosen in a normal distribution with a user-defined standard deviation  $sd\_H$  corresponding to an expected number of alleles and rate of heterozygosity (See Fig. S6). For example, when  $sd\_H = 1$ , the expected heterozygosity is around 70% and the number of possible alleles is around 10.



**Fig.S6:** Heterozygosity and number of alleles corresponding to the parameterization of  $sd\_H$  in the GUI, using 10000 random values in a normal distribution

**Table S2.** Default values at initialization.

<i>Variable names</i>	<i>Definition</i>	<i>Default value</i>
<b>INIT.</b>		
Wd	Working directory	
nb_agent	Number of individuals per cell	100
in_X	X Coordinate of individuals	10
in_Y	Y Coordinate of individuals	4
everywhere	To specify that individuals are located in every cells	False
From_fstat_file	To specify if individuals are created according to an input file	False
Sd_H	Standard deviation of the normal distribution defining the number of alleles and heterozygosity among the population	1
biall?	Switch to choose biallelic loci with either “0” or “1” instead of a normal distribution	False
<b>LANDSCAPE</b>		
Land_management	Type of land management	None
Land_time_step	Time steps for land management	1000
habitat_type_file	Name of the file containing habitat types	“habitat_type.txt”
habitat_barrier_file	Name of the file containing habitat barriers	“habitat_barrier.txt”
habitat_resource_file	Name of the file containing habitat resources	“habitat_resource.txt”
<b>INDIVIDUALS</b>		
num_microsat	Number of microsatellites loci (or SNPs)	10
num_locus_per_habitat	Number of loci under selection per habitat type	1
proba_dispersion	Dispersion rate	0.5
nb_move	Number of movements	1
r_growth	Growth parameter of the logistic growth submodel	0.5
var_coef_selec_vs_xx	Selection coefficient against the deleterious genotype	0.5
var_dom_degree_a	Degree of dominance of the deleterious allele	0.5
mutation_rateN	Mutation rate	10 <sup>-4</sup>
<b>OUTPUTS</b>		
output_freq	Outputs frequency (every $x$ time steps)	100
var_num_generations	Number of generations	100
var_num_repetitions	Number of repetitions	1
CSV	Comma-separated values output file	True
GENEPOP	Genepop output file	False
FSTAT	Fstat output file	False
ARLEQUIN	Arlequin output file	False
STRUCTURE	Structure output file	False
GENELAND	Geneland output file	False
sample_pop?	Sampling of individuals	True
n_points	Number of points per habitat type	20
n_ind	Number of individuals per point sampled	25
SRS	To specify if the sampling is made with (WR) or without (WOR) replacement	WR

## 6. Input data

There are no input data.

## 7. Submodels

Some submodels descriptions include a “validation” and “verification” subheading. Here, “validation” refers to the justification of submodels used, according to common practices and literature in the relevant fields (*i.e.* rationale; see section III for a validation of SimAdapt as defined by Rykiel *et al.* 1996). “Verification” refers here to the correctness of the model formalism, *i.e.* a correct implementation of the conceptual model (Rykiel *et al.* 1996; Parker *et al.* 2003).

### 7.1. Individuals related submodels

#### 7.1.1. Reproduction (main\_reproduction)

##### *Description*

The reproduction submodel is based on the assumption that all individuals located in a given cell have a equi-probability of mating, weighted with individual selective values, and all the gametes produced constitute a gamete pool. Then transmission of genetic characteristics follows Mendel inheritance laws. The number of descendants is defined by the logistic growth submodel according to fitness traits depending on the habitat type (`habitat_type`). The parameters of the logistic growth function are the number of individuals in a given cell before reproduction ( $N_0$ ), the user-defined intrinsic growth rate ( $r = r\_growth$  in the GUI), and the carrying capacity defined by the habitat resources ( $K = habitat\_resource$ ). The population genetics submodel is detailed in the main text and not copied in this supplementary material. Mutations occur only in microsatellites loci with a probability of  $10^{-4}$  which can be modified through the GUI.

##### *Validation*

The population genetics submodel is based on Hartl and Clark (2007), Wade *et al.* (2001) and Trajstman (1973). We assumed a stepwise mutation model as defined by Hamilton (2009, p 169).

##### *Verification*

In the code the selective values  $w_i$  of each individual is stored in a list (`allwi` in procedure `main_reproduction`). Then two gametes are chosen in the gamete pool with a probability based on  $w_i$  so that selection is conserved. The selection of two gametes is reproduced  $N$  times for the creation of the next generation. This submodel has been checked by verifying that selection was operational over a large number of individuals with a broad range of selection coefficients. To verify that Mendel inheritance laws were adequately reproduced by our model, we simulated more than 1000 first generations and checked the allele frequencies of 10 loci (allele frequency of 0.49995, CI95% = [0.4952743 ; 0.5046257]).

##### *Pseudo-algorithm*

```
FOR all cells {  
    IF at least 2 individuals {  
        define number of offspring Nt1 from Nt0  
        REPEAT Nt1 times {
```

```

        choice into gamete pool (probability wi)
        create descendant with Mendel inheritance laws
    }
}
remove old generation of individuals
}

```

### 7.1.2. Survival (**main\_reproduction**)

#### *Description*

Each cell in the grid is characterized by a carrying capacity defining the maximum number of individuals that a cell can contain (variable `habitat_resource`). The number of individuals is controlled by a logistic growth submodel.

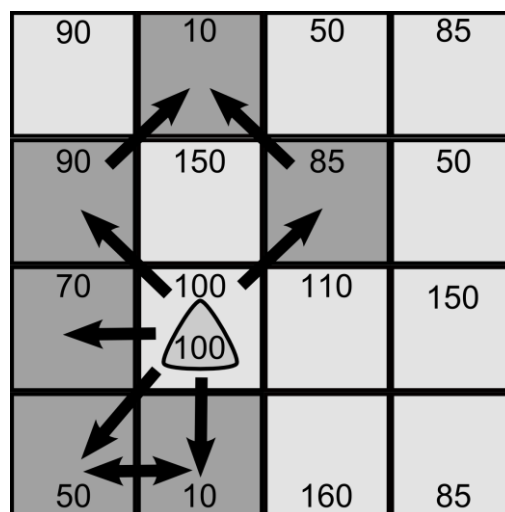
#### *Verification*

This submodel can be checked at any moment through the GUI using the population dynamics plot which monitors the number of individuals as a function of time.

### 7.1.3. Dispersion (**main\_dispersion**)

#### *Description*

Each individual can move from one cell to another located in its Moore neighborhood (8 nearest neighbors). Each cell is characterized by its resistance (patch variable `habitat_barrier`). Lowest values of resistance represent easy to cross cells and highest values impermeable cells. Each individual is characterized by its dispersion capabilities (individual variable `cap_move`) and its maximum dispersal distance (individual variable `nb_move`). An individual with higher dispersion capabilities could move to more cells within its Moore neighborhood (see Fig. S7). The decision whether to move to another cell is based on a fixed probability (*i.e.* migration rate, default probability of 0.5). The destination cell is chosen randomly among potential destinations. This feature can be easily changed in the code to reproduce more realistic behaviors such as movements toward the cells maximizing fitness for agents having a perception of their environment, or movements in a given direction whenever possible for migratory agents.



**Fig. S7.** Dispersal representation of individuals across the landscape. An individual is located in initial cell with a dispersion capability of 100 and a maximal dispersal distance of 2 cells. Potential destinations are represented in dark grey and movements with arrows.

#### *Validation*

We assumed that individuals do not have a perception of the suitability of neighboring cells and disperse randomly into the landscape.

#### *Verification*

This submodel has been checked by verifying that the number of individuals selected for dispersal corresponded to the rate of dispersion, by verifying that the potential destination cells were adequate to dispersal capabilities, that the individuals moving were properly removed from the buffer list, and finally that individuals were properly relocated to chosen destinations.

#### *Pseudo-algorithm*

```
FOR all individuals {
  WHILE individual can move {
    list of potential destinations
    move to a random cell among potential destinations
    update of dispersal capacity according to movement
  }
}
```

## **7.2. Initialization related submodels**

### **7.2.1. Setting up the list of alleles for neutral and under selection loci**

Each allele for microsatellites loci is chosen in a normal distribution of standard deviation `sd_H` to obtain a given heterozygosity and number of possible alleles. Alternatively, biallelic loci can be chosen to represent SNPs by switching on `biall?` to “on” (allele “0” and “1” randomly chosen). Alleles at loci under selection are chosen randomly at initialization.

### **7.2.3. Setting up the individuals**

Identification numbers are incremented during the reproduction process and for the simulations up to 100 generations, and then reinitiated at each generation. Number of alleles and heterozygosity are presented in “input” section.

## **7.3. Landscape related submodels**

### **7.3.1. Landscape generation**

External files:

- “*habitat\_type.txt*” for the landscape habitat type,
- “*habitat\_resource.txt*” for the landscape carrying capacity and
- “*habitat\_barrier.txt*” for the landscape resistance.

In case files contain errors, a default value is attributed to each layer.



### 7.3.2. Landscape management

Landscape can change over time due to human actions. The simulation model includes a landscape management submodel that reproduces various scenarios.

Alternatives are:

- |                  |  |
|------------------|--|
| - "none"         | no landscape management                                      |
| - "random"       | random changes over time                                     |
| - "neighbor4"    | changes to one of the 4 neighbors (Von Neumann neighborhood) |
| - "neighbor8"    | changes to one of the 8 neighbors (Moore neighborhood)       |
| - "transition"   | changes to one of the existing habitat type                  |
| - "user-defined" | to let the user program his own submodel                     |

### 7.4. Output

#### *Description*

In order to perform population genetics analysis, the simulation model produces output files for the most-used population genetics software (see main text). The sampling method is based on random location without replacement (a warning message is provided when the number of sampling locations `n_points` is higher than the number of locations in the grid: in such cases the location is not sampled twice, resulting in a lower number of sampled individuals in output files) and random individuals (with or without replacement).

#### *Validation*

The choice of software was realized following Excoffier and Heckel 2006.

#### *Verification*

Each output has been checked using the corresponding software.

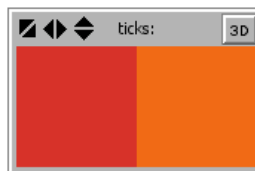
### III. VALIDATION

Validation refers here to the comparison between output results of SimAdapt simulations with those of population genetics models and other existing software results (see Rykiel 1996 for a discussion on validation). In order to validate SimAdapt software, we performed simplified simulations that could be compared with existing analytically known results. In a first section is addressed the evolution of alleles frequencies at loci under selection compared to population genetic theory. The second section addresses the evolution of heterozygosities at neutral markers in a panmictic population compared to previous work of Landguth *et al.* 2010b. The third section addresses the evolution of the fixation index (F) using one of the most widely model of the gene flow among subpopulation, the island model, compared to existing theory revised by Hamilton in 2009 (specifically pp. 135-141).

#### 1. Validation by comparing SimAdapt simulation results including selection with population genetic theoretical expectations

##### 1.1. SimAdapt

To compare SimAdapt with population genetics analytical results including selection, we performed two sets of simulations on a two-cell landscape and two habitat types (see Figure S8 for parameterization). The carrying capacity was of 1000 individuals per grid cell in the first set of simulations and 100 individuals per grid cell in the second set of simulations. These simulations were repeated 30 times for different values of migration rate (0, 0.02, and 0.5), during 100 generations. We reported the frequencies of the allele “1” at the loci under selection associated with the *b1* habitat type (on the left in figure S8) and associated with the *b2* habitat type (on the right in figure S8), for individuals located in the *b1* habitat type. The initial frequency of the allele “1” at the locus associated with the *b1* habitat type was set to 0.5 and to 0.8 for the other locus.



**Fig. S8.** Representation of the simulation on SimAdapt GUI with two grid cells and two habitat types (*b1* and *b2*).

##### 1.2. Population genetic analytical model with selection and infinite population size

The analytical model considers two infinite populations (*b1* and *b2* habitat types) between which individuals migrate at rate  $m$  (0, 0.02 and 0.5). The frequency of the allele “1” at loci under selection on the *b1* habitat type was simulated according to an analytical population genetic model, according to the equation:

$$\left\{ \begin{aligned} P_{t+1,bl,j} &= (1-m) \frac{P_{t,bl,j}^2 + P_{t,bl,j}(1-P_{t,bl,j})(1-hs)}{P_{t,bl,j}^2 + 2P_{t,bl,j}(1-P_{t,bl,j})(1-hs) + (1-P_{t,bl,j})^2(1-s)} + m \frac{P_{t,b2,j}^2(1-s) + P_{t,b2,j}(1-P_{t,b2,j})(1-hs)}{P_{t,b2,j}^2(1-s) + 2P_{t,b2,j}(1-P_{t,b2,j})(1-hs) + (1-P_{t,b2,j})^2} \\ &\text{and} \\ P_{t+1,b2,j} &= m \frac{P_{t,bl,j}^2 + P_{t,bl,j}(1-P_{t,bl,j})(1-hs)}{P_{t,bl,j}^2 + 2P_{t,bl,j}(1-P_{t,bl,j})(1-hs) + (1-P_{t,bl,j})^2(1-s)} + (1-m) \frac{P_{t,b2,j}^2(1-s) + P_{t,b2,j}(1-P_{t,b2,j})(1-hs)}{P_{t,b2,j}^2(1-s) + 2P_{t,b2,j}(1-P_{t,b2,j})(1-hs) + (1-P_{t,b2,j})^2} \end{aligned} \right\} \quad (\text{eq. 1})$$

With:

$p_{t+1}(b1,j)$ : frequency of allele “1” at locus  $j$  in the  $b1$  habitat type

$p_{t+1}(b2,j)$ : frequency of allele “1” at locus  $j$  in the  $b2$  habitat type

$m$ : migration rate

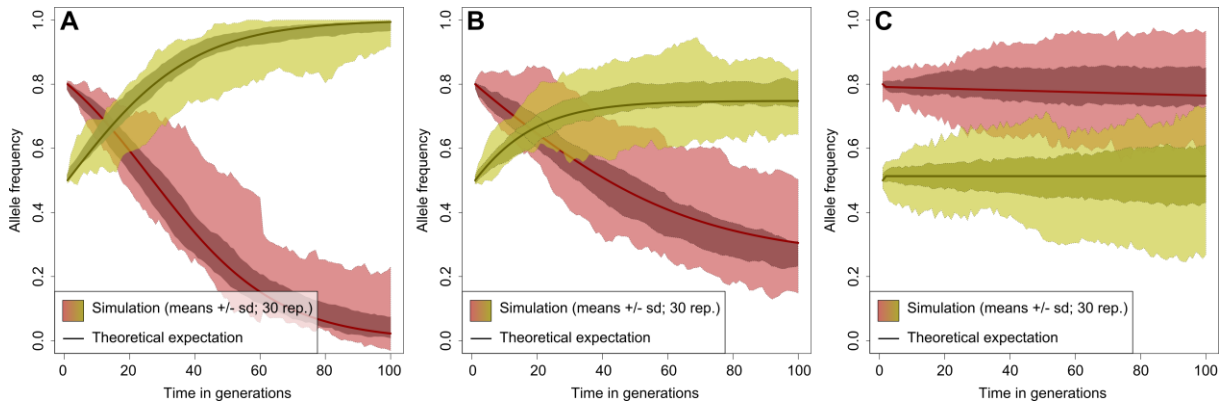
$h$  and  $s$ : selection coefficients

$j$ : locus under selection associated with the  $b1$  habitat type

The initial frequency of the allele “1” at the locus under selection associated with the  $b1$  habitat type was set to 0.5 and the frequency of the allele “1” at the locus under selection associated with the  $b2$  habitat type was set to 0.8 (see figure S9).

### 1.3. Results

SimAdapt simulated results were concordant with theoretical expectations (see figure S9). When migration rate produce a population close to panmixia (*i.e.* migration rate of 0.5 between  $b1$  and  $b2$  habitat type, see figure S9 C), the allele “1” associated with the  $b1$  habitat type are not selected through time, while the selection of this allele is clearly observed with low migration rate (Figure S9 B), and drastically with a null migration rate resulting in the fixation of the allele “1” (figure S9 A). The stochasticity observed is attributed to the number of individuals (high genetic drift with 100 individuals per habitat type, lowered with 1000 individuals per habitat type), and to the stochastic nature of dispersion (see details section of the ODD protocol).



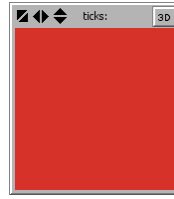
**Fig. S9.** Frequencies of the allele “1” at the loci under selection associated with the  $b1$  habitat type (in yellow) and the  $b2$  habitat type (in red) for individuals located in the  $b1$  habitat type, as a function of time in generations (initial frequency values of 0.5 and 0.8, respectively). SimAdapt simulation results (yellow and red areas) correspond to the mean frequency values of 30 repetitions +/- standard deviation for populations of 100 individuals (light area) and 1000 individuals (dark area) per habitat type. Theoretical expectations (solid yellow and red

lines) are based on a population genetic analytical model (see eq. 1). Migration rates were parameterized to A: 0; B: 0.02; and C: 0.5.

## 2. Validation by comparing SimAdapt simulation results without selection with population genetic theoretical expectations in a single panmictic population

### 2.1. SimAdapt

To compare SimAdapt with population genetics analytical results without selection, we simulated a landscape made of one grid cell (panmictic population, see figure S10), with 1248 individuals and 10 microsatellite loci, during 500 generations, as in Landguth *et al.* (2010b).



**Fig. S10.** Representation of the simulation on SimAdapt GUI with one grid cell.

### 2.2. Population genetic analytical model without selection and finite population size

Theoretical heterozygosities for neutral markers in a single panmictic population were performed according to Landguth *et al.* (2010b), *i.e.*:

$$H_t = \left(1 - \frac{1}{2N_e + 1}\right)^t H_o(0) \quad (\text{eq. 2})$$

Where

$$N_e = \frac{4 N_M N_F}{N_M + N_F} \quad (\text{eq.3})$$

$H_t$ : theoretical rate of decay after  $t$  generations

$H_o$ : proportion of heterozygous individuals at each time step

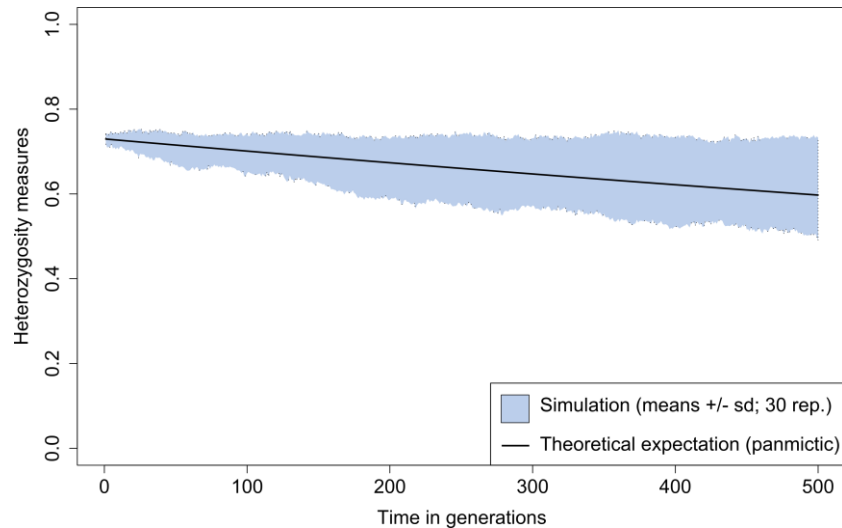
$H_o(0)$ : initial heterozygosity

$N_e$ : effective population size

$N_M$  and  $N_F$ : total number of males and females, respectively

### 2.3. Results

When comparing the simulations with expected results, the simulations fitted theoretical expectations (see figure S11), providing an additional validation of SimAdapt model for neutral markers in a panmictic population.

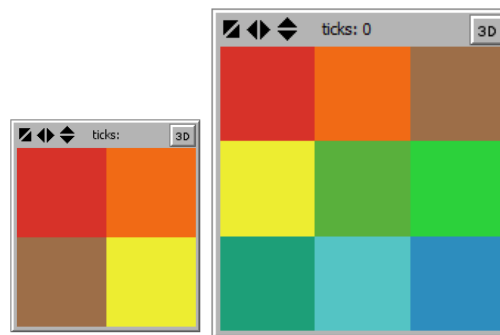


**Fig. S11.** Heterozygosities at neutral loci as a function of time in generations for all individuals (carrying capacity of 1248 individuals). SimAdapt simulation results (light blue area) correspond to the mean heterozygosity of 30 repetitions  $\pm$  standard deviation. Theoretical expectations (solid black line) for a panmictic population were based on Landguth *et al.* 2010b.

### 3. Validation by comparing SimAdapt simulation results without selection with population genetic theoretical expectations in an island model

#### 3.1. SimAdapt

An infinite island model assumes an infinite number of identical subpopulations (demes), where the probability of exchanging migrants between subpopulations is equal. Due to the spatially explicit nature of SimAdapt, we performed simulations of a finite island model made of 4 subpopulations (closed boundaries) and 9 subpopulations (open boundaries, *i.e.* toroidal grid) with 1000 individuals and 10 neutral loci (see figure S12). As in section 1 and 2, simulations were repeated 30 times. Two sets of simulations were performed: in the first set, migration was null and in the second set, migration was parameterized to 0.02. Simulation results (FSTAT input files) at each generation were analyzed using the “pegas” R package (Paradis 2010) to compute fixation index values for the 10 neutral loci (loci under selection disabled for these sets of simulations).



**Fig. S12.** Representation of the simulation on SimAdapt GUI with four grid cells (closed boundaries) and nine grid cells (toroidal).

### 3.2. Population genetic analytical model

To compare SimAdapt with existing theory, we used the infinite island model as described by Hamilton 2009, where the fixation index (F) as a function of time in generations is:

$$F_t = \frac{1}{2N_e} (1-m)^2 + \left(1 - \frac{1}{2N_e}\right) F_{t-1} (1-m)^2 \quad (\text{eq.4})$$

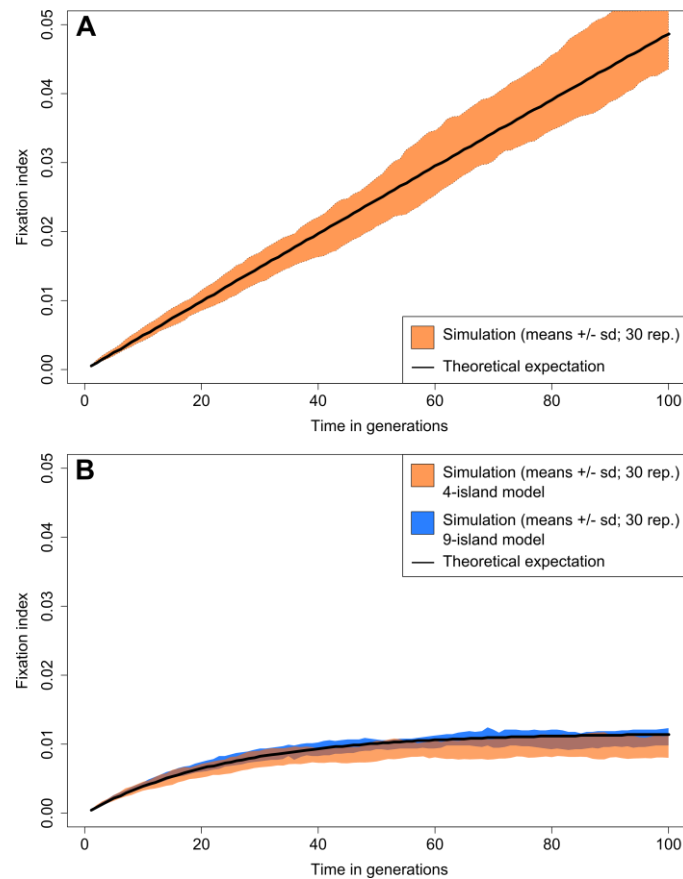
$m$ : migration rate

$t$ : time in generations

$N_e$ : effective population size

### 3.3. Results

When the migration rate is null (see figure S13 A), the expected results were concordant with SimAdapt simulations, while we observed a slight deviation in the case of a positive migration rate in the 4-island model, reduced in the case of the 9-island model (see figure S13 B). This slight deviation is explained by the finite number of subpopulations in SimAdapt simulations (see Hamilton 2009 p138 for a discussion on genetic drift in finite island models).



**Fig. S13.** Fixation index as a function of time in generations in an island model. SimAdapt simulation results correspond to the mean fixation index values of 30 repetitions +/- standard deviation for an island model with four subpopulations (orange area) and nine subpopulations (blue area) of 1000 individuals and 10 neutral loci. In A: the migration rate was null and in B: the migration rate between each population was 0.02. Theoretical expectations (solid black lines) were based on Hamilton 2009.

#### IV. EXAMPLE FILE “HABITAT\_TYPE.TXT”

This file is made of three columns: the abscissa (x from 0 to 19, *i.e.* 20 grid cells), the ordinate (y from 0 to 4, *i.e.* 5 grid cells) and the habitat type (from 1 to 2, *i.e.* 2 habitat types). This type of file can be easily generated from an ascii file (exported from any Georeferenced Information System program), using the “ascii\_to\_simadapt.r” script provided on the SimAdapt webpage of the OpemABM consortium (<http://openabm.org/models/>).

0	0	1
1	0	1
2	0	1
3	0	1
4	0	1
5	0	1
6	0	1
7	0	1
8	0	1
9	0	1
10	0	2
11	0	2
12	0	2
13	0	2
14	0	2
15	0	2
16	0	2
17	0	2
18	0	2
19	0	2
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	2
11	1	2
12	1	2
13	1	2
14	1	2
15	1	2
16	1	2
17	1	2
18	1	2
19	1	2
0	2	1
1	2	1
2	2	1
3	2	1
4	2	1
5	2	1
6	2	1
7	2	1
8	2	1

9	2	1
10	2	2
11	2	2
12	2	2
13	2	2
14	2	2
15	2	2
16	2	2
17	2	2
18	2	2
19	2	2
0	3	1
1	3	1
2	3	1
3	3	1
4	3	1
5	3	1
6	3	1
7	3	1
8	3	1
9	3	1
10	3	2
11	3	2
12	3	2
13	3	2
14	3	2
15	3	2
16	3	2
17	3	2
18	3	2
19	3	2
0	4	1
1	4	1
2	4	1
3	4	1
4	4	1
5	4	1
6	4	1
7	4	1
8	4	1
9	4	1
10	4	2
11	4	2
12	4	2
13	4	2
14	4	2
15	4	2
16	4	2
17	4	2
18	4	2
19	4	2



## REFERENCES

- Bruggeman D.J., Jones M. L., Scribner K.T., Lupi, F., 2009. Relating Tradable Credits for Biodiversity to Sustainability Criteria at a Landscape-Scale. *Landscape Ecology* 24, 775-790.
- Bruggeman D.J., Wiegand T., Fernández N., 2010. The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). *Molecular Ecology* 19, 3679-3691.
- Duenez-Guzman E.A., Mavarez J., Vose DM.D., Gavrillets S., 2009. Case studies and mathematical models of ecological speciation. 4. Hybrid speciation in butterflies in a jungle. *Evolution* 63, 2611-2626.
- Excoffier L and Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics* 7, 745-758.
- Gavrillets S., Vose A., 2005. Dynamic patterns of adaptive radiation. *PNAS* 102, 18040-18045.
- Gavrillets S., Vose A., Barluenga M., Salzburger W., Meyer A., 2007. Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Molecular Ecology* 16, 2893-2909.
- Grimm V., Berger U., Bastiansen F., Eliassen S., Ginot V., Giske J., Goss-Custard J., Grand T., Heinz S., Huse G., Huth A., Jepsen JU., Jørgensen C., Mooij WM., Müller B., Pe'er G., Piou C., Railsback SF., Robbins AM., Robbins MM., Rossmanith E., Rüger N., Strand E., Souissi S., Stillman RA., Vabø R., Visser U., DeAngelis DL., 2006. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* 198, 115-126.
- Grimm V., Berger U., DeAngelis DL., Polhill G., Giske J., Railsback SF., 2010. The ODD protocol: a review and first update. *Ecological Modelling* 221, 2760-2768
- Guillaume F, Rougemont J (2006) Nemo: an evolutionary and population genetics programming framework. *Bioinformatics*, 22, 2556-2557.
- Hamilton M.B., 2009. Population genetics. Eds John Wiley & Sons, 407pp.
- Hartl, D.L., Clark, A.G. (2007) *Principles of population genetics* (eds Sinauer Associates), pp. 565.
- Holderegger R., Wagner H.H., 2006. A brief guide to landscape genetics. *Landscape Ecology* 21, 793-796.
- Jaquiéry J., Broquet T., Hirzel A.H., Yearsley J., Perrin N., 2011. Inferring landscape effects on dispersal from genetic distances : how far can we go ? *Molecular Ecology* 20, 692-705.
- Landguth E.L., Cushman S.A., 2010b. CDPOP : A spatially explicit cost distance population genetics program. *Molecular Ecology Resources* 10, 156,161.
- Landguth E.L., Cushman S.A., Murphy M.A., Luikarts G., 2010a. Relationships between migration rates and landscape resistance assessed using individual-based simulations. *Molecular Ecology Resources* 10, 854,862.
- Landguth E.L., Cushman S.A., Schwartz M.K., McKelvey K.S., Murphy M., Luikart G., 2010c. Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology* 19, 4179-4191.
- Lawson D., Jensen H.J., 2006. The species-area relationship and evolution. *Journal of Theoretical Biology* 241, 590-600.
- Le Page C., Becu N., Bommel P., Bousquet F., 2012. Participatory Agent-Based Simulation for Renewable Resource Management: The Role of the Cormas Simulation Platform to

- Nurture a Community of Practice. *Journal of Artificial Societies and Social Simulation*, 15(1) 10.
- Manel S., Schwartz M.K., Luikart G., Taberlet P., 2003. Landscape genetics : combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18, 189-197.
- Manel S., Segelbacher G., 2009. Perspectives and challenges in landscape genetics. *Molecular Ecology* 18, 1821,1822.
- Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNEMO: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, 24, 1552-1553.
- Paradis E (2010) pegas: an {R} package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419-420.
- Parker D.C., Manson S.M., Janssen M.A., Hoffmann M.J., Deadman P., 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers*, 93, 314-337.
- Peng B, Kimmel M (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, 21, 3686-3687.
- Phillips P.D., Thompson I.S., Silva J.N.M., van Gardingen P.R., Degen B., 2004. Scaling up models of tree competition for tropical forest population genetics simulation. *Ecological Modelling* 180, 419-434.
- Rykiel E.J., 1996. Testing ecological models: the meaning of validation. *Ecological Modelling*, 90, 229-244.
- Saladin S., Littlejohn M., 2003. A spatially explicit individual-based model of reinforcement in hybrid zones. *Evolution* 57, 962-970.
- Thiele J.C., Kurth W., Grimm V., 2012. RNetLogo: an R package for running and exploring individual-based models implemented in NetLogo. *Methods in Ecology and Evolution*, 3(3): 480-483.
- Trajstman A.C., 1973 The necessity of the poisson distribution for the equivalence of some random mating models. *Mathematical Biosciences* 17, 1-10.
- Wade MJ, Winther RG, Agrawal AF, Goodnight CJ (2001) Alternative definitions of epistasis: dependence and interaction. *TRENDS in Ecology and Evolution* 9, 498-504.
- Wilensky U (1999) NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.
- Zurell D., Berger U., Cabral J. S., Jeltsch F., Meynard C. N., Münkemüller T., Nehrbass N., Pagel J., Reineking B., Schröder B., Grimm V., 2010. The virtual ecologist approach: simulating data and observers. *Oikos*, 119: 622–635. doi: 10.1111/j.1600-0706.2009.18284.x