APPLICATION

SimAdapt: an individual-based genetic model for simulating landscape management impacts on populations

François Rebaudo[1,2]*, **Arnaud Le Rouzic**[2], **Stéphane Dupas**[1,2], **Jean-François Silvain**[1,2], **Myriam Harry**[2] and **Olivier Dangles**[1,2,3]

[1]*Biodiversité et évolution des complexes plantes-insectes ravageurs-antagonistes, IRD-BEI-UR072, 91198 Gif-sur-Yvette Cedex, France;* [2]*Laboratoire Évolution Génome et Spéciation, CNRS-LEGS-UPR9034, Université Paris-Sud 11, 91198 Gif-sur-Yvette Cedex, France; and* [3]*Facultad de Ciencias Naturales y Biológicas, Pontificia Universidad Católica del Ecuador, Quito, Ecuador*

Summary

1. Simulation models are essential tools in landscape genetics to study how genetic processes are affected by landscape heterogeneity. However, there is still a need to develop different simulation approaches in landscape genetics, so that users may dispose of additional programs to explore further the impact of land-use and land-cover changes on population genetics.

2. We developed a spatially explicit, individual-based, forward-time, landscape-genetic simulation model combined with a landscape cellular automaton to represent evolutionary processes of adaptation and population dynamics in changing landscapes, using the NetLogo environment.

3. This simulation model represents a unique tool for scientists and scholars looking for a practical and pedagogical framework to explore both empirical and theoretical situations.

Key-words: adaptation, ecology, forward-time, land-use, NETLOGO, spatially explicit

Introduction

Evolutionary ecologists have increasingly recognized the importance of local adaptive genetic variation to predict population response to changing environmental conditions (Manel & Segelbacher 2009). During the last 5 years, the field of landscape genetics has developed numerous simulation modelling approaches to better understand population adaptation in natural conditions (Balkenhol & Landguth 2011), in particular through backward-time simulation models allowing landscape-genetic inferences (e.g. IBDSIM, Leblois, Estoup & Rousset 2008; SPLATCHE2, Ray *et al.* 2010; SIMSSD, Legendre, Borcard & Peres-Neto 2005; see Hoban, Bertorelle & Gaggiotti 2012 for review). In parallel, population-genetic, forward-time, simulation models (e.g. CDPOP, Landguth & Cushman 2010b; SIMUPOP, Peng & Kimmel 2005; NEMO, Guillaume & Rougemont 2006; QUANTINEMO, Neuenschwander, Hospital & Goudet 2008; see Hoban, Bertorelle & Gaggiotti 2012 for review) have been developed to explore and test empirical hypotheses *in silico*. However, in a context where equilibrium is the exception, especially in landscapes under anthropogenic disturbance (Epperson *et al.* 2010), the possibility of modifying landscape characteristics in most simulation software

remains rudimentary or only accessible to skilled programmers (e.g. CDPOP, SPLATCHE2). If the identification of appropriate spatial and temporal scale is still a challenge (Balkenhol *et al.* 2009), a flexible spatially explicit landscape submodel for landscape genetics is required for realistic simulations to be studied and compared (Balkenhol & Landguth 2011).

To contribute filling this gap, we developed a landscape-genetic, forward-time, simulation model in which users can easily and explicitly represent landscape changes over space and time, thereby simulating the genetic dynamics of individuals in landscapes with different management scenarios (see Table 1 for a comparison of SIMADAPT with SPLATCHE2 and CDPOP). The main innovations of our model are the dispersal submodel (behavioural model allowing complex movements, see Supporting information) and the natural selection submodel including any number of loci under selection (most traits are known to be the result of multiple loci). Our software uniqueness relies on the coupling of two submodels: a population-genetic submodel taking advantage of an individual-based approach (see Landguth *et al.* 2010), and a cellular automaton, landscape submodel. The later allows the integration and exploration of complex landscape scenarios, empowered by the large literature existing on land-use and land-cover changes (e.g. Parker *et al.* 2003; Crespo-Pérez *et al.* 2011). The simulation model is user-friendly to target a broad audience of scientists [graphical user interface (GUI), see Fig. 1], cross-platform

Table 1. Features of cdpop (v1.2.08beta), simadapt and splatche2 (v2.01)

Software	Operating System and interface	Model	Data type	Dispersal model	Population dynamics	Natural selection	Landscape dynamics	Sensitivity analysis module	Output files
CDPOP (Landguth & Cushman 2010b)	Linux, Windows (GUI and console)	Individual-based model	Allele frequencies, mtDNA option	Dispersal distance functions (sex-specific)	Fluctuating population size: carrying capacity, birth and death rates.	One or two biallelic loci (% offspring mortality per habitat type)	Input files loaded at defined generations: landscape resistance (mating and dispersal)	Option to perform multiple runs and explore parameters	Input files for Structure, GenePop, Genelex. *csv* raw files and build-in analyses
SIMADAPT	Any OS with Java virtual machine (GUI, console and R package)	Individual-based model	STRs, SNPs	Migration rate and dispersal distance (behavioural submodel)	Logistic growth model	Any number of biallelic loci per habitat type (s and h coefficients)	Independent submodel (cellular automaton): natural selection, carrying capacity, landscape resistance with build-in flexible scenarios of land-use and land-cover changes.	Dedicated software within NetLogo to explore parameters with multiple runs (behaviour space)	Input files for Structure, FSTAT, Arlequin, GenePop, Geneland. *csv* raw files and virtual sampling model
SPLATCHE2 (Ray *et al.* 2010)	Linux (console), Windows (GUI and console)	Coalescent model	DNA sequences, SNPs, STRs, RFLPs	Mutation rate	Logistic growth model	No	Input files loaded at defined generations: carrying capacity, landscape resistance	Option to perform multiple runs	Input files for Arlequin, Nexus. Build-in *ascii/bmp* files (analyses)

GUI, graphical user interface.

and freely available on the openabm website (http://www.open abm.org/models/). It is written in the high-level language netlogo (Wilensky 1999), which confers easily extensions and navigation through the annotated source code (see Railsback & Grimm 2012 for an introduction to individual-based models and netlogo). In addition, it generates output files directly readable by most used and up-to-date population genetics computer programs (see Excoffier & Heckel 2006 for a selected list), and can be run from the free statistics software r (R Development Core Team 2012; see Thiele, Kurth & Grimm 2012, and Appendix S1, Supporting information). Here, we describe the model and the corresponding implementation, and exemplify its potential uses.

Components of SimAdapt

SimAdapt simulates the evolution of both neutral and adaptive genotypes of diploid, sexually reproducing individuals introduced in a landscape. It accounts for the transmission of genes according to Mendelian inheritance laws, dispersal and adaptation to local conditions. This model combines a landscape cellular automaton submodel (CA) and an individual-based, spatially explicit submodel (IBM). Each time step in the model corresponds to one generation for individuals (with non-overlapping generations).

CELLULAR AUTOMATON

The CA, represented by an area with closed boundaries, and which dimensions are user-defined, includes a three-layer georeferenced information system (allowing text files integration from classic GIS programs) to characterize the landscape: (i) the available resources (carrying capacity for individuals), (ii) the landscape resistance (permeable or semi-permeable barriers for migration) and (iii) the type of habitat (for natural selection). The landscape characteristics can vary over space and time with five different predefined scenarios: L1) no changes in the type of habitat; L2) random changes; L3), L4) changes to an adjacent habitat type (using the Moore -8 nearest neighbours- and the von Neumann-4 nearest neighbours-neighbourhood (von Neumann 1948), respectively); and L5) transition towards a homogeneous landscape consisting of one habitat type. With basic programming skills, additional scenarios can be easily defined with the possibility to set habitat type changes over time through the GUI (see Fig. S2 in Supporting information).

INDIVIDUAL-BASED MODEL

The IBM represents the individuals living in the landscape (parameters of spatial dynamics and local adaptation). Initially, individuals are located either at a given set of coordinates or homogeneously over the landscape. They are characterized by a dispersal capability affecting the spatial pattern (see Supporting information for a detailed description and Epperson *et al.* 2010 for a discussion). Briefly, individual can move from one cell to another located in the eight nearest neighbours up to a maximal dispersal distance (i.e. individuals potentially
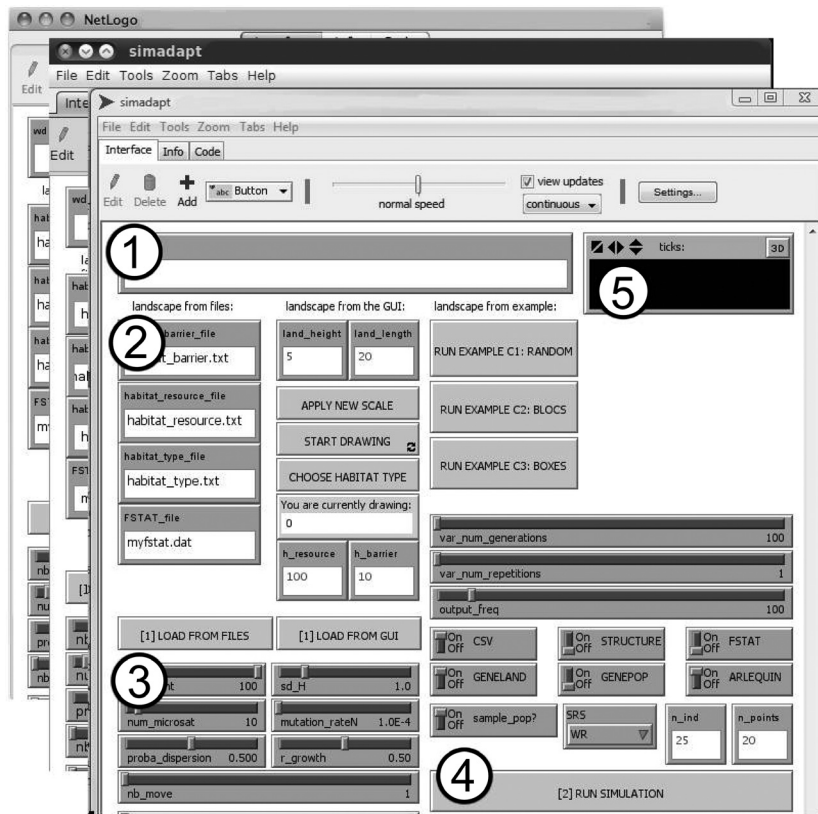
Fig. 1. Graphical user interface from SIMADAPT in different operating systems. SIMADAPT has a simple layout of options divided into areas such as: (1) working directory; (2) landscape features; (3) genetic features; (4) simulation features; and (5) visualization.

move across multiple cells). Each cell is characterized by its resistance (from permeable to impermeable cells), and each individual is characterized by its dispersion capabilities. The decision whether to move to another cell is based on a fixed probability (i.e. migration rate), and the destination cell is chosen randomly among potential destinations (see Fig. 2).

Markers under selection

Reproductive fitness of individuals is the result of the habitat-dependent selection. The population-genetic submodel within the IBM assumes one or several bi-allelic loci under selection per habitat type. Alleles are either counter selected (represented by '0') or selected (represented by '1') in each habitat type. Being adapted to a given habitat is considered advantageous while the individual is in this habitat, but harmful in others habitats. This allows the program to simulate multiple loci which is important to analyse selection and speciation (Epperson *et al.* 2010). Following the notation used by Hartl & Clark (2007), the value s is the selection coefficient against the homozygous G_{00} genotype (where the indices refer to the first and second allele at the locus k under selection) and h is the degree of dominance of the 0 allele.

Consequently, at the locus k associated with habitat type $b1$, while $b2$ representing another habitat type, the relative fitness w of genotypes G_{11}, G_{10} and G_{00}, for individuals located in habitat type $b1$ and $b2$, respectively, are:
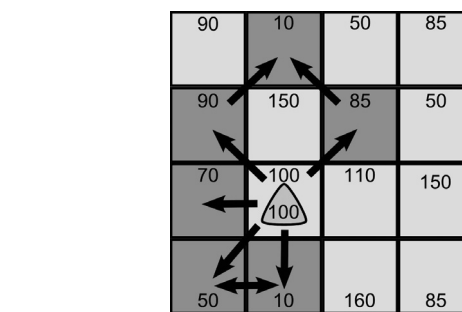

Fig. 2. Dispersal representation of individuals across the landscape. An individual is located in initial cell with a dispersion capability of 100 and a maximal dispersal distance of two cells. Potential destinations are represented in dark grey and movements with arrows.

$$w_{G_{11},k,b1} = 1 \quad \text{and} \quad w_{G_{11},k,b2} = 1 - s \qquad \text{eqn 1a}$$

$$w_{G_{10},k,b1} = 1 - sh \quad \text{and} \quad w_{G_{10},k,b2} = 1 - sh \qquad \text{eqn 1b}$$

$$w_{G_{00},k,b1} = 1 - s \quad \text{and} \quad w_{G_{00},k,b2} = 1 \qquad \text{eqn 1c}$$

A multiplicative model with no epistasis is assumed, and the selective value of a given genotype is the product of the selective values at each locus (Wade *et al.* 2001):

$$W_{I,b1} = \prod_{k=1}^{K} w_{G_I,k,b1} = 1 \qquad \text{eqn 2}$$

with I an individual on habitat type $b1$.

For each location (grid cell) in the landscape, individuals (the parental generation) produce N offsprings, with N following a logistic growth function based on the number of individuals in the considered location, a user-defined intrinsic growth rate and a carrying capacity defined by the habitat resources. For each offspring, two parents are drawn randomly among parental individuals at the same location, proportionally to their fitness, and a gamete is generated from each parent (allowing the genealogy to be reconstructed from output files). The genetic transmission follows the Mendelian inheritance laws, assuming free recombination between loci.

Neutral markers

In addition to the set of loci involved in adaptation, n neutral independent loci (microsatellites typically) can be considered. Z alleles can be present simultaneously in the population at each locus (at the initialization step, alleles are chosen in a normal distribution with a user-defined standard deviation conditioning on the number of alleles and the expected heterozygosity, see Fig. S6 in Appendix S1, Supporting information; alternatively, users can choose biallelic loci with '0' and '1' to represent SNPs). Mutational events, at a rate of μ mutations per allele per generation, replace allele z by allele $z + 1$ or by allele $z-1$, according to a classical stepwise mutation model. As for loci under selection, Mendelian inheritance and free recombination between loci are assumed for the neutral loci.

MODEL OUTPUT

The characteristics of each individual are stored in a file including habitat type and coordinates during time (see Table S2 in Appendix S1, Supporting information for an exhaustive list of characteristics). To allow direct comparison between simulated and experimental data through specialized programs, the simulation model includes an empirical sampling module. A number of recollection points per habitat type are defined by the user, each of them with a given number of sampled individuals (see Zurell *et al.* 2010 for a discussion on sampling in simulation models). The output file contains the microsatellite genotypes of sampled individuals in a format that can be processed by most used population genetics software including GENEPOP v4.1 (Rousset 2008), ARLEQUIN v3.1 and v3.5 (Excoffier, Laval & Schneider 2005), STRUCURE v2.3.3 (Pritchard, Stephens & Donnelly 2000), GENELAND v3.3 (Guillot, Mortier & Estoup 2005) and FSTAT (Goudet 1995).

IMPLEMENTATION

A complete description and documented verification of the simulation model following the Overview, Design concepts, Details (ODD) protocol for describing individual- and agent-based models (Grimm *et al.* 2006, 2010) and the executable code are provided as Supporting information (Appendices S1 and S2), along with SimAdapt validation. Basic validation tests were performed to make sure that the software results are consistent with classical population-genetic models, including the comparison of (i) alleles frequencies at loci under selection with an analytical model of infinite population size, (ii) heterozygosity measures at neutral markers with theoretical expectations in a single panmictic population and (iii) fixation index between populations as a function of time with theoretical expectations in an island model (see Supporting information). The code is documented and structured to be modified and extended with basic programming skills. If a reasonable simulation size should not include more than 10,000 individuals in a 100 grid cells landscape with 100 neutral markers and 100 loci under selection in most computers, our simulation model has no imposed limits and allows any users to facilitate the process of getting the first insights to explore landscape-genetic complexity and to visualize local adaptation.

Study example

We present here an example to provide a brief overview of the simulation model functionalities and coupling with existing programs. We simulated 100 individuals with $n = 10$ microsatellites loci introduced in one grid cell of a landscape composed of two habitat types with one locus under selection per habitat type (habitat types $b1$ and $b2$, see Fig. 3 and Supporting information for parameterization). Simulations were repeated 100 times for 100 generations, and output sampled for each habitat type. Three different habitat configurations were tested (see Fig. 3a): C1) random location of habitat types with landscape scenario L2; C2) blocks of habitat types with scenario L1 (i.e. no changes); and C3) isolated habitat types with scenario L1. These three basic scenarios represented C1) a rapidly changing heterogeneous landscape, where individuals have to adapt constantly, C2) a spatial transition between two distinct habitats (e.g. field/forest; urban area/natural area) and C3) a more heterogeneous landscape arising from habitat fragmentation. They were provided to give a basic understanding of the program functionality and performance, but also to illustrate the ability to integrate complex landscapes changing over space and/or time. In Fig. 3b, we mapped the distribution of allelic frequencies at a locus under selection using all individuals. Using output files containing microsatellites data, we generated basic genetic analyses with ARLEQUIN, with the assumption that individuals located in different habitat types correspond to distinct population. We then used GENELAND to visualize the correlation between habitat types and population structure (see Fig. 3c). Simulation examples can be reproduced through the GUI (<40 s per simulation in a laptop using MS Windows Vista, 4Go RAM, CPU 3.06 GHz).

Conclusion

Thanks to its modularity, this software represents a unique tool to explore the interactions between gene flow, population
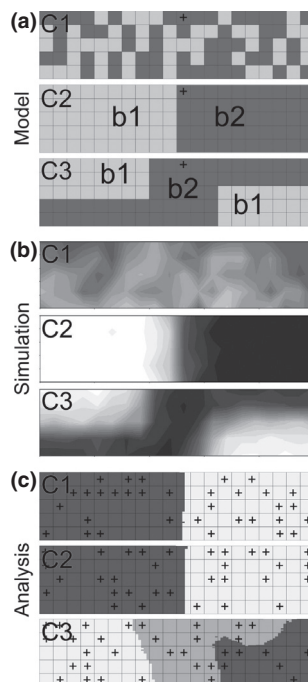
Fig. 3. Example of SIMADAPT use with (a) three landscape configurations build in the graphical user interface (GUI) referred as C1, C2 and C3 with b1 in light grey and b2 in dark grey representing two different habitat types and the cross the initial locations of individuals; (b) simulation results showing the frequency of one allele at a locus under selection of the first habitat type after 100 generations, ranging from zero in black to one in white; (c) population assignment analysis based on neutral markers using GENELAND, with each color representing a different population and each cross a point sampled.

dynamics, selection and landscape management, to link landscapes and adaptive genetic variation (Parisod & Holderegger 2012). Relying on a coupled individual-based model and cellular automaton, this framework allows the integration of complex patterns including spatially and temporally explicit landscapes described at an accurate level. This singular coupling favours the identification of land-use and land-cover change management scenarios driving population structure. From theoretical cases to empirical studies, it should facilitate our understanding of landscape genetics and represents a promising tool for scientists and scholars willing to explore ecosystems subject to anthropogenic disturbance and social-ecological systems complexity.

Users are encouraged to visit and submit comments to the SIMADAPT web page hosted by the OPENABM consortium (http://www.openabm.org/models/).

Acknowledgements

References

Balkenhol, N. & Landguth, E. (2011) Simulation modelling in landscape genetics: on the need to go further. *Molecular Ecology*, **20**, 667–670.

Balkenhol, N., Gugerli, F., Cushman, S., Waits, L., Coulon, A., Arntzen, J.W., Holderegger, R. & Wagner, H. (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology*, **24**, 455–463.

Crespo-Pérez, V., Rebaudo, F., Silvain, J.F. & Dangles, O. (2011) Modeling invasive species spread in complex landscapes: the case of potato moth in Ecuador. *Landscape Ecology*, **26**, 1447–1461.

Epperson, B.K., Mcrae, B.H., Scribner, K., Cushman, S.A., Rosenberg, M.S., Fortin, M.J., James, P.M.A., Murphy, M., Manel, S., Legendre, P. & Dale, M.R.T. (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology*, **19**, 3549–3564.

Excoffier, L. & Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, **7**, 745–758.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Goudet, J. (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Vabø, R., Visser, U. & DeAngelis, D.L. (2006) A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, **198**, 115–126.

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, G., Giske, J. & Railsback, S.F. (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.

Guillot, G., Mortier, F. & Estoup, A. (2005) Geneland: a program for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Hartl, D.L. & Clark, A.G. (2007) *Principles of Population Genetics*, Fourth edition pp. 565. Sinauer Associates, Sunderland, MA, USA.

Hoban, S., Bertorelle, G. & Gaggiotti, O.E. (2012) Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics*, **13**, 110–122.

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. (2010) Relationships between migration rates and landscape resistance assessed using individual-based simulations. *Molecular Ecology Resources*, **10**, 854–862.

Leblois, R., Estoup, A. & Rousset, F. (2008) IBDSim: a computer program to simulate genotypic data under isolation by distance. *Molecular Ecology Resources*, **9**, 107–109.

Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.

Manel, S. & Segelbacher, G. (2009) Perspectives and challenges in landscape genetics. *Molecular Ecology*, **18**, 1821–1822.

Neuenschwander, S., Hospital, 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.

von Neumann, J. (1948) The general and logical theory of automata. *Cerebral Mechanisms in Behavior* (ed. L.A. Jeffress), pp. 1–41. Wiley, New York.

Parisod, C. & Holderegger, R. (2012) Adaptive landscape genetics: pitfalls and benefits. *Molecular Ecology Resources*, **21**, 3644–3646.

Parker, D.C., Manson, S.M., Janseen, M.A., Hoffman, 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.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. URL http://www.r-project.org/

Railsback, S.F. & Grimm, V. (2012) *Agent-based and Individual-based Modeling: A Practical Introduction*. Princeton University Press, Princeton, New Jersey, USA.

Ray, N., Currat, M., Foll, M. & Excoffier, L. (2010) SPLATCHE2: a spatially-explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, **26**, 2993–2994.

Rousset, F. (2008) GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

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**, 480–483.

Wade, M.J., Winther, R.G., Agrawal, A.F. & Goodnight, C.J. (2001) Alternative definitions of epistasis: dependence and interaction. *Trends in Ecology and Evolution*, **9**, 498–504.

Wilensky, U. (1999) *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, Illinois, USA. URL: http://ccl.northwestern.edu/netlogo/.

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.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Documentation and extended presentation of the model following the ODD protocol together with validation and verification documentation (latest version available with the model at http://www.openabm.org/models/).

Appendix S2. Simulation model (code included) using NETLOGO multi-agent programmable modelling environment (latest version available at http://www.openabm.org/models/).