HAL open science

Sensitivity analysis of a crop metapopulation model

Baptiste Rouger, Isabelle Goldringer, Pierre M Barbillon, Anne Miramon, Abdel Kader Naino Jika, Mathieu Thomas

To cite this version:

Baptiste Rouger, Isabelle Goldringer, Pierre M Barbillon, Anne Miramon, Abdel Kader Naino Jika, et al.. Sensitivity analysis of a crop metapopulation model. Ecological Modelling, 2023, 475, pp.110174. 10.1016/j.ecolmodel.2022.110174 . hal-03887218

HAL Id: hal-03887218

https://hal.inrae.fr/hal-03887218v1

Submitted on 6 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Ecological Modelling 475 (2023) 110174

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Sensitivity analysis of a crop metapopulation model

Baptiste Rouger [a,b,*], Isabelle Goldringer [a], Pierre Barbillon [c], Anne Miramon [a], Abdel Kader Naino Jika [a,d], Mathieu Thomas [e,f]

[a] Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE – Le Moulon, 91190, Gif-sur-Yvette, France
[b] Université de Paris, France
[c] Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, 75005, Paris, France
[d] Bioversity International, 00057, Rome, Italy
[e] CIRAD, UMR AGAP, F-34398 Montpellier, France
[f] AGAP, Univ Montpellier, CIRAD, INRAE, Montpellier SupAgro, Montpellier, France

ARTICLE INFO

Keywords:
Genetic diversity
Dynamic management
Seed network
Agent-based model
Agroecology

ABSTRACT

CropMetaPop is a new simulation tool to model the genetic evolution of crop diversity under on-farm dynamic management. Under this type of conservation and use of varieties, seeds are resown and exchanged between farmers and the set of connected populations is described as a crop metapopulation. CropMetaPop is therefore at the interface of genetic and social processes. We used sensitivity analyses to check the behaviour of the model and to identify which parameters and range of values for them induce the most variability in the outputs. CropMetaPop was found to behave as expected. Depending on the type of locus studied (neutral or selected), the parameters related to drift or selection were those that induced the most variability in the outputs. Colonisation, migration, and network topology parameters were less influential. Looking at the detailed results will help setting the parameters to relevant values in the future utilisation of the model.

1. Introduction

In small-scale farming systems, farmers save their seeds from one year to the next and often exchange seeds for a variety of reasons including the search for better production, seed loss, or even curiosity (Pautasso et al., 2013). Seed circulation among farmers has been shown to greatly impact the genetic diversity of the plant populations they grow in their fields (Fuentes et al., 2012; Louette et al., 1997; Pressoir and Berthaud, 2004).

Those crop populations are linked by past or future seed transfers and are subject to extinction. For these reasons, they can be considered as a particular case of metapopulation, called crop metapopulation which is submitted to farmers management (van Heerwaarden et al., 2010).

Farmers' seed management practices and the functioning of these cultivated metapopulations are at the interface between social and ecological processes and understanding the impacts of farmers' practices and organisation on the dynamics of crop populations diversity is a complex issue that seems unlikely to be tackled through experimental (Pautasso et al., 2013) or analytical approaches only.

Indeed, setting up experiments aimed at comparing the impact of different management methods would be far too cumbersome, too long (e.g. to observe allele fixation time (Kimura, 1980)) and therefore too costly.

A modelling approach is therefore well suited to such studies, which allows cheap and fast results though they might lack accuracy. Several metapopulation models exist, such as QuantiNemo 2 (Neuenschwander et al., 2019) which is considered the most comprehensive one available. The extensivity of this model can address four features specific to crop metapopulations as describes by van Heerwaarden et al. (2010). These are (1) the large number of offspring per plant, (2) the non-random seed circulation, (3) the low rate of seed circulation, and (4) the rare bottlenecks during recolonisation events.

Despite this, several key features are misrepresented in QuantiNemo 2. Extinction occurs after the regulation of adults (see Fig. 1(a)), whereas it seems more likely to occur during the seed storage phase in crop metapopulations, after seed production and before the seedling or adult phases, i.e., before regulation. The order of events during a life cycle has been shown in the literature to influence the evolution of local adaptive polymorphism (Ravigne et al., 2004; Massol and Débarre, 2015). Although the authors do not explicitly address extinctions, we assume that such a change is likely to have the same kind of impact on the results. For this reason, it seems important to represent extinction at the most likely phase of the cycle.

Furthermore, QuantiNemo 2 does not distinguish between migration and colonisation, whereas the two processes allow to represent

* Correspondence to: Equipe DEAP, GQE - Le Moulon, 12 rue 128, 91190 Gif-sur-Yvette, France,
E-mail address: baptiste.rouger@disroot.org (B. Rouger).

https://doi.org/10.1016/j.ecolmodel.2022.110174
Received 30 March 2022; Received in revised form 19 September 2022; Accepted 12 October 2022
Available online 3 November 2022
0304-3800/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

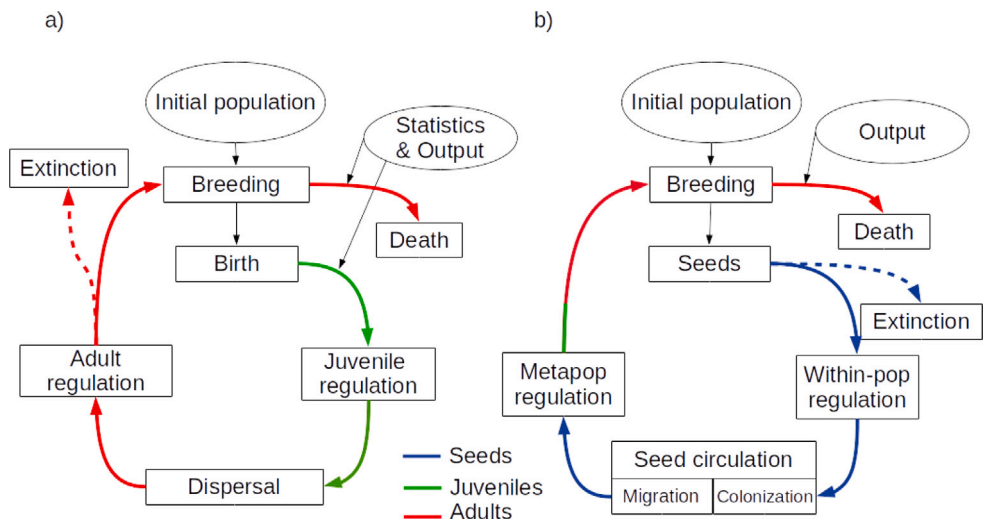

Fig. 1. Comparison between life cycle used in: (a) QuantiNemo 2 (adapted from the QuantiNemo 2 user manual (Neuenschwander et al., 2019, p. 34)); (b) CropMetaPop. The main difference between QuantiNemo 2 and CropMetaPop (CMP) population life cycles is that most of the events in CMP happen during the seed phase, while QuantiNemo 2 is divided in juvenile and adult phases. In CMP, during within-population regulation the number of seeds produced by a population is reduced to the carrying capacity according to individual plant fitness. Metapopulation regulation only occurs in case of colonisation or migration. In the case of colonisation, metapopulation regulation reduces the number of seeds in the recipient population to its carrying capacity according to the average fitness of the donor populations and their seed availability. In the case of migration, metapopulation regulation occurs only on the part of the local seeds to be replaced by migrant seeds. QuantiNemo 2 performs summary statistics, while CMP only outputs mono- or multilocus genotype data and seed circulation history.

two coexisting situations that are classically encountered by farmers who produce their own seeds. Migration allows to represent situations quite frequently encountered in subsistence farming systems where farmers need to supplement their own seed stocks to sow their entire available field. In this context, they used to acquiring seed locally within their village from relatives, neighbours or friends through gift, exchange or purchase (Boster, 1986; Wencélius et al., 2016; Thomas and Caillon, 2016; Violon et al., 2016). Colonisation coupled with extinction represents more extreme but less frequent situations in which farmers no longer have seed stocks after a poor harvest due to climatic hazards or because of problems with seed storage (Fenzi et al., 2021; Violon et al., 2016). It has been described that in this type of situation, people change their seed supply network by going to obtain seed from relatives or acquaintances outside the village, in some case travelling very long distances (Fenzi et al., 2021; Violon et al., 2016; Bellon et al., 2011; Song et al., 2019). The simultaneous consideration of these two social processes into account offers the possibility of representing relatively accurately situations commonly encountered in agriculture.

Finally, an other critical feature (van Heerwaarden et al., 2010) is to define independently the seed transfer rate and the level of seed replacement between every pair of populations, because in crop management migration is episodic rather than continuous and involves enough seeds to limit the bottleneck.

All these deviations from classical metapopulation models such as those implemented in QuantiNemo 2 prompted us to develop CropMetaPop, a model of cultivated metapopulations in the continuity of the work that was conducted by van Heerwaarden et al. (2010) to investigate the impact of seed exchanges and farmers' management practices.

CropMetaPop[1] is an agent-based model where agents are individuals represented by plants or seeds, which was specifically designed to model cultivated populations, and the gene flows resulting from seed circulation due to the social interactions between farmers (Labeyrie et al., 2016; Calvet-Mir et al., 2012; McGuire, 2008).

CropMetaPop simulates the evolution of individuals represented as multilocus genomes organised into populations, in turn organised into a metapopulation. It aims at simulating the evolution of seed lots within the context of seed exchange following the social network of the farmers.

The life cycle of individuals in CropMetaPop is described in Fig. 1, and presents the comparison between the life cycles of CropMetaPop and QuantiNemo 2.

It allows to model various genetic situations: low or high genetic diversity, weak or strong population differentiation, for instance. Moreover, different parameters can be specified to best represent the crop of interest: mating system, number of individuals per population, etc. One can also finely tune the seed movement parameters for migration and for colonisation such that the probability of seed exchange is defined between all pairs of populations.

Due to the large number of parameters in the model, performing a sensitivity analysis (SA in the following) (Saltelli et al., 2019; Burgers et al., 2010) of the model was needed to better understand its behaviour and to ensure that it performs as expected. Sensitivity analyses also provide a better understanding of which input parameters are responsible for the variability of the model's outputs (Lurette et al., 2009; Lamboni et al., 2011). Finally, one can use SAs for a better targeting of a realistic range of parameters to be used in the model.

The two objectives of this paper are (i) to introduce the CropMetaPop model and software and (ii) to conduct several SAs of the CropMetaPop model in order to check that the model behaves as expected and to assess the relative importance of the parameters in the model based on the variability they induce in the outputs.

2. Materials and methods

This section begins with a description of the CropMetaPop model. It continues with the presentation of the main steps to be taken when performing a sensitivity analysis before introducing the 6 sensitivity analyses performed in this paper.

2.1. The CropMetaPop model

2.1.1. Technical features of CropMetaPop

CropMetaPop can be seen as a wrapper library for simuPOP specifically designed to easily simulate crop metapopulations. simuPOP (Peng

[1] The software is available at https://forgemia.inra.fr/cropmetapop/cropmetapop/-/wikis/home

and Kimmel, 2005; Peng and Amos, 2008) is a python library for the simulation of stochastic individual-based model accounting for mutation, genetic drift, migration and selection and for demographic processes such as extinction and colonisation. CropMetaPop provides a useful and practical addition to simuPOP by integrating specific characteristics of crop metapopulations, such as in particular the possibility to generate or import various connection matrices to link the fields together. It also integrates modules to create random networks with specific characteristics. It is therefore designed to represent accurately seed transfer processes with specific parameters. CropMetaPop is a console program written in Python3 using an object-oriented approach. It allows running simulations without knowledge about coding by just filling a text based simulation file and it can be run on any computer platform.

Tests restricted to one or sometimes two evolutionary forces at a time made it possible to verify that the model was able to find the theoretical expectations (data not shown) and thus to validate the broad lines of its operation.

2.1.2. General features

In the context of seed management by farmers' organisations, a crop metapopulation may correspond, for instance, to different versions, here referred to as populations, of the same variety of a considered species. CropMetaPop relies on the following life cycle: breeding, extinction, within-population regulation, colonisation, migration and meta-population regulation. The model works with discrete and non-overlapping generations, which makes it more suitable for representing annual or short generation time species.

Depending on the life cycle stage, each individual is either a plant in the field or a seed that may circulate among populations. In the following, the populations are equivalent to the demes as well as the fields to the patches.

2.1.2.1. Demographic features of CropMetaPop. CropMetaPop accounts for several social features of the farmers' networks (number of farmers involved and number of populations per farmer) and farming practices (sowing density and field size) by simulating a finite number of populations, each composed of a finite number of crop plants corresponding to the demographic size. Each population is cultivated in a field characterised by the maximum number of plants that can be grown there, called carrying capacity. The demographic size may vary from one generation to the following one according to the number of offspring produced per individual (called fecundity). The demographic size of a crop population can grow up to the carrying capacity of the field in which the population is cultivated. Each population is subject to extinction, colonisation or migration in each generation. It should be noted that pollen flows between fields are neglected in CropMetaPop because we consider that farmer practices limit such phenomenon. They take care to grow different varieties at sufficient distances in the case of mainly self-pollinating species and will generally only grow one variety, under isolated conditions, in the case of a cross-pollinating species (e.g. pearl millet (Naino Jika et al., 2017), maize or onion).

In the context of crop metapopulation, extinction may correspond either to difficulties in maintaining the population (e.g. climatic or pest disasters) or to the choice of replacing the population with a potentially more interesting one. Colonisation and migration will be defined in the following section.

2.1.2.2. Genetic features of CropMetaPop. To account for a critical biological feature of crops, each crop metapopulation is characterised by a selfing rate ranging from 0 (open-pollination) to 1 (self-pollination). Each diploid individual plant (i) in a given field (j) consists in a finite set of genetically linked or independent loci. Each locus is biallelic or multi-allelic. Mutation rate is defined for each locus and mutation may occur at each reproduction event. Each locus may correspond to a neutral marker or to an adaptive marker located in a gene associated to a genetic value. The sum of the genetic values of all adaptive markers provides the individual genetic value of a quantitative trait (G_i) under selection. Natural selection due to the local pedo-climatic conditions or to particular farming practices is accounted for by applying selection for a local optimum defined for each field. Depending on the local field optimum (Opt_j), the individual fitness ($W_{i,j}$) is assumed to have a classical normal shape (Johnson and Barton, 2005) centred on the local optimum:

$$W_{i,j} = e^{-\frac{1}{2}\times(G_i - Opt_j)^2}$$

with i an individual plant and j a given field.

Similar or different optima can be assigned to the different fields in order to compute plant fitness and apply various patterns of selection to local populations. During the breeding step, the parents of each offspring are randomly selected proportionality to their fitness when genetic values and fields' optima are defined.

Within-population regulation is applied after seed production to adjust potential excess of offspring to the carrying capacity of each population.

2.1.3. Special features

In CropMetaPop, the focus has been on describing seed circulation practices in order to represent a diversity of socially established rules that underpin the circulation of seed among farmers.

Two different processes are considered to model seed circulation: colonisation and migration. Colonisation arises after the extinction of a local population while migration happens when farmers mix intentionally or unintentionally their own seed lot with foreign ones. Unlike other metapopulation simulation softwares, CropMetaPop can model both colonisation and migration in the same simulation, with the possibility to use two different social networks, if necessary, because we consider that farmers can solicit two different social networks depending on the context. In both cases, seed circulation is a stochastic process and seed supply can come from one or from several non-empty fields when farmers are socially connected.

The amount of seed in circulation is defined at the level of metapopulation regulation on the basis of donor seed supply and receiver seed demand.

For instance, one farmer can give seeds to neighbours according to his or her stock and receive seeds from them depending on his or her needs. In addition for migration, it is possible to define the rate of local seed replacement which can be different from the migration rate. This specific feature is necessary to take into account the fact that migration rarely occurs but with a potentially strong impact in terms of genetic composition on the receiving population when the seed replacement rate is huge.

2.2. Sensitivity analysis approach

CropMetaPop is a stochastic simulation model that allows to adjust the values of more than 10 parameters. Predicting the behaviour and results of simulations can be very complicated in some situations. To better understand the general behaviour of the model, we propose to use SAs.

Performing a SA of a model first requires defining the input parameters and output variables of the model to be analysed. It is then necessary to run the model for a large number of combinations of input parameter values. For each combination, values are obtained for the different output variables of the model.

SA relies on variance analysis to attribute the variability of the outputs of the model to one of the parameters changing in the simulations (Cariboni et al., 2007; Burgers et al., 2010; Carpani et al., 2012).

In this paper, we chose to vary the main parameters that control biological features and evolutionary forces of the crop metapopulations, i.e. the selfing rate to characterise the mating system, the population size for genetic drift, the mutation rate, the number of locus under

selection and local selection optima, extinction and colonisation rate, migration rate and network topology for colonisation and migration.

In order to study in depth the evolutionary processes of particular interest to us, we have divided the work into six SAs. This allowed us to better describe the impact of each parameter in each SA.

The parameters chosen to vary in these SAs were:

carr_cap carrying capacity: the number of individuals in the population

percentSelf percent of selfing: the proportion of individuals in the population that are autogamous

mutRate mutation rate of each locus

nbLocSel number of locus of the genome that are under selection

optimum optimums of the demes: the set of optimums to which the genotypic value of the individuals are compared to compute the fitness of each individual in the deme

colRate colonisation rate between two populations

extinction extinction rate of the populations

network network topology type. See Section 2.2.3 for the tested graph topologies

nbEdges number of edges in the network, i.e. the number of social relationships in the metapopulation

migRate migration rate (i.e. probability) of seed lots between two populations

migReplace proportion of replaced individuals in the destination population during migration

The other parameters of the simulations have been set to fixed values. The most important ones are displayed in Table 1.

The number of generations (**generations**) to achieve the SAs was set at 30, which represents situations that can be observed in reality. Beyond that, it seems unlikely that the social networks are the same, for example. Therefore, exploring the behaviour of the model over a longer period of time does not seem to be a priority with respect to the conditions of use of the model. The number of replicates (**replicates**) was set to 10 to capture some of the stochasticity of the model while limiting the simulation time of the SAs. The number of populations (**nb_pop**) was set at 100, which corresponds to a reasonable number of people involved in a collective action of crop diversity management. The number of alleles (**nb_allele**) was set to two to represent SNPs. The fecundity parameter (**fecondity**) was set to 4 to produce enough offspring to allow for migration events while limiting simulation time. The number of neutral markers (defined as the difference between the total number of markers (**nb_marker**) and the number of markers associated with a fitness value) was set to 10 to be able to follow the evolution of several loci simultaneously while limiting the simulation time.

To assess the impact of the parameters variation, genetic diversity indices (see Section 2.2.4) were computed on genotypic data.

The protocol was the same for the six SAs. It consisted in the following steps:

1. Creation of the design of experiments with the correct number of parameter (Section 2.2.1)
2. Creation of the simulations setting files and of the launcher files for the simulations (Section 2.2.2)
3. Launch of the simulations
4. Analysis of the genetic monolocus data (Section 2.2.4) launched with the scripts generated in Section 2.2.2
5. Gathering of all the files in big files by indicator
6. SA of the resulting files (Section 2.2.5)
7. The graphs interpretation of the SAs (Section 2.2.6)

Table 1
Common parameters to all simulations.

Parameter	Use of the parameter	Value
generations	Number of generations in the simulation	30
replicates	Number of replicates in the simulation	10
nb_pop	Number of populations in the metapopulation	100
nb_allele	Number of alleles for all loci	2
fecundity	Number of seeds produced by individual	4
no associated parameter	Number of neutral markers	10

2.2.1. *Creation of the plan of simulation*

In order to avoid running thousands of simulations, a fractional factorial design was applied using the R package "planor" (Kobilinsky, 2005; Kobilinsky et al., 2017, 2019) for each of the six SAs of CropMetaPop.

This package allowed us to run less than 500 simulations where 20 000 would have been needed with a full factorial design. It requires to set each parameter chosen to an equal number of level, and allows to determine up to second order interactions between the parameters, i.e. interaction of one parameter with one other. It returns a design of experiments composed of a list of levels combinations. This design of experiments was exported to a comma separated value formatted file and used as an input for the scripts generating the setting files for the simulations and the analysis to run.

In this paper, we chose to use three levels for each parameter. We considered it allows to correctly – yet concisely – explore the variation ranges of the parameters of the model. We chose, for every parameter, a low realistic value, a high realistic value, and an intermediary value.

2.2.2. *Creation of the simulations*

The setting file generator creates one configuration file for each replicate and for each of the levels combination in the design of experiments. It associates to the set of levels described in the design of experiment, the correct corresponding values to each parameter. It also writes two more files: one that contains the set of simulations to run on the computer cluster (the simulation launcher file), and the other that contains the corresponding analyses to run after the simulations (the analysis launcher file).

The ten replicates of every parameter combination are split in separate simulations in order to maximise the variability of both the network topology used, and the genetic diversity at initialisation. These replicates are then grouped together to perform the analysis of monolocus data.

The simulations are initialised genetically by sampling, for each locus of each individual, a random allele between the two available at each locus. This process can induce an imbalance in the frequencies of each allele, especially in small size populations. This can also induce linkage disequilibrium between markers.

2.2.3. *Network topologies tested*

We used three contrasted topologies chosen to be representative of the reality of the systems we model. The Erdős–Renyi network topology (Erdős and Rényi, 1959) in which the connections between the populations are randomly and statistically equally distributed, was designed to represent a perfectly distributed and horizontal network. We then used the community network topology (Nowicki and Snijders, 2001; Girvan and Newman, 2002), in which the connections between the populations are more frequent within a community, to represent the collective management of seeds in different farmers' organisations.

Last, we used the Barabási (Barabási and Albert, 1999) topology network, in which only a few populations are connected to many other

populations. This was chosen to represent centralisation around a few major actors in the seed system. The densities have been chosen to represent networks with low, medium and almost-maximal densities.

For the SAs that focused on migration or colonisation, a library of networks was generated before the simulations to maximise the network topology stochasticity. A total of 900 networks was generated in the library. A network with corresponding parameters was randomly selected in the library and written in the configuration file.

2.2.4. Monolocus data analysis

For each SA, after the simulations, the data of each replicate that have the same parameter combination, and that are in separate folders, were gathered in the same file. All the files, *i.e.* all simulations, were then analysed by a Python script to compute two genetic and one demographic indicators for each generation, including the average within-population expected genetic diversity (expected heterozygosity) (Hs (Nei, 1987)), the average genetic differentiation (Gst (Weir and Cockerham, 1984; Takahata and Nei, 1984)) and the survival rate (Surv) of the populations.

$$Hs = \frac{\sum_{s=1}^{S} \frac{\sum_{l=1}^{L}\left(1-\sum_{i=1}^{I} p_{i,l,s}^2\right)}{L}}{S}$$

$$Gst = \frac{\sum_{s=1}^{S} \frac{Ht_s - Hs_s}{Ht_s}}{s}$$

with

$$Ht = \frac{\sum_{l=1}^{L}\sum_{i=1}^{I} p_i^2}{l}$$

and

$$Surv = 1 - \frac{\text{number of non-empty populations}}{\text{total number of populations}}$$

with I the number of alleles, L the number of loci and S the number of populations.

This analysis produced as many result files as there were combinations of parameters and indicators. These files were then merged by indicator to produce the final result files of the simulations. These indicator files contain the average indicator values over the ten replicates.

2.2.5. Sensitivity analysis

For each SA, the resulting files were then analysed for each indicator with the R "multisensi" package (Bidot et al., 2018; Lamboni et al., 2009, 2011). It allows to evaluate the share of every parameter in the variability of the output along with time, determining up to second order interactions of parameters. It takes as inputs a merged indicator data file and the design of experiments created by "planor" and returns detailed sensitivity indices, as well as several graphs, including one presenting the evolution of the share of every parameter to the variability of the output indicator of the model along time (such as in Figs. 2 and 4).

2.2.6. Graphs interpretation

For each SA, the lower part of the graph represents the main sensitivity indices for each input parameter normalised between 0 and 1. The bigger the width of the ribbon of the parameter, the more the parameter is important to explain the variability of the model output at the given generation. For clarity, interactions of parameters are merged and represented in a unique ribbon. The residual represents the part of the variability of the model that cannot be explained by any parameter or second order interaction.

The upper part of the graph represents the distribution of the output of the model. The dark heavy line represents the mean of the data, the grey area represents the quartiles of the data. The blue dotted lines cover 90% of the data, and the red dotted lines the extreme data. One can consider that, the wider the distribution, the more significant the effects we detect.

Table 2
Parameter values for ISOLATED POPULATIONS.

Parameter	Neutral SA	Markers	Selected SA	Markers
carr_cap	40, 100, 1000			
percentSelf	0, 0.5, 0.95			
mutRate	10^{-5}, 10^{-4}, 10^{-3}			
nbLocSel	0, 1, 10		1, 5, 10	
optimum	1s only, 1/2 0s and 1/2 1s, gradient between 0 and 1			

The Figs. 2(c), 3(c), 4(c), 5(c) and 7 present the total sensitivity indices (i.e. the sum of the sensitivity indices of each parameter and of its interactions with other parameters) for the last generation of all SAs. As the last generation sensitivity indices is the sum of interactions attributed to two parameters, the sum of all shares can be superior to 1.

2.3. A six-sensitivity analyses system

2.3.1. Description

Six SAs were performed to identify the parameters that induce the most variability in the results according to the evolutionary forces considered. These SAs differ according to two criteria: if and how the populations are connected (isolated, connected by colonisation, connected by migration) and the type of markers studied (neutral or under selection). Their names follow a pattern describing first the connection of populations (ISOLATED POPULATIONS, COLONISATION–EXTINCTION, or MIGRATION) and then the type of markers studied (NEUTRAL MARKERS or SELECTED MARKERS).

Note that the results of two SAs are not comparable since they are not based on the same set of input value combinations.

The values of the parameters for each SAs are presented in Tables 2–4. For each of the parameter chosen, we defined three levels, which were fixed to span over most of the actual variation ranges of the parameters.

Mutation rates were chosen to cover the range of values that can be found *in vivo* (Raquin et al., 2008; Schoen and Schultz, 2019).

The number of loci under selection were chosen to represent cases with no selection and with selection. When selection was present in the simulations, we chose to represent cases with a monogenic traits, or traits composed by 5 or 10 loci. Even though these numbers are quite low compared to the number of loci associated with traits *in vivo* (e.g. for wheat height (Zanke et al., 2014)), we were able to show expected dynamics for selection, and spared computation time.

Other parameters were chosen to cover a wide range of realistic values for every parameter, with one intermediary value, in order to have contrasted effects with the parameter range:

carr_cap ranged from 40 (vegetable-like size of population) to 1000 to represent cereal populations. Those populations can be a lot bigger, but genetic drift is estimated to be fairly reduced from 1000 individuals (Nei et al., 1975)

percentSelf ranged from 0 (purely allogamous individuals) to 0.95 (mostly autogamous individuals), to represent the most species possible

optimum were chosen to represent contrasted situations of selection where:

- individuals are all in the same environment
- individuals are in two contrasted environments
- individuals are in environments all different one from another

Table 3
Parameter values for SAs Colonisation– Extinction.

Parameter	Neutral SA	Markers	Selected SA	Markers
carr_cap				
percentSelf				
mutRate	Identical to Table 2		Identical to Table 2	
nbLocSel				
optimum				
network	Erdős–Renyi, Community, Barabási			
colRate	0.01, 0.1, 0.25			
nbEdges	$E_{\max} \times 0.04$, $E_{\max} \times 0.08$, $E_{\max} \times 0.16$			
extinction	0.01, 0.1, 0.25			

With $E_{\max} = \frac{nbPop \times (nbPop-1)}{2}$.

Table 4
Parameter values for SAs C Migration.

Parameter	Neutral SA	Markers	Selected SA	Markers
carr_cap				
percentSelf				
mutRate	Identical to Table 2		Identical to Table 2	
nbLocSel				
optimum				
network	Erdős–Renyi, Community, Barabási			
migRate	0.01, 0.1, 0.25			
nbEdges	$E_{\max} \times 0.04$, $E_{\max} \times 0.08$, $E_{\max} \times 0.16$			
migReplace	0.05, 0.2, 0.5			

With $E_{\max} = \frac{nbPop \times (nbPop-1)}{2}$.

network More information about network topologies can be found in Section 2.2.3

colRate and migRate were chosen to represent situations with very few to a lot of colonisation/migration, i.e. one event in a hundred generations to one event every four generations per population.

extinction were chosen to represent situations with very few to a lot of extinction, i.e. one event in a hundred generations to one event every four generations per population.

migReplace were chosen to represent situations where only a hundredth of the population is replaced, to half of the population is replaced by migration events

nbEdges More information about network topologies can be found in Section 2.2.3

2.3.2. Hypotheses

Qualitative hypotheses regarding the nature of the results are proposed for each of the six SA, based on the main principles of population genetics.

2.3.2.1. Hypotheses for isolated populations neutral markers.

Here, as we analyse the genetic diversity indicators at neutral markers that are not linked to selected genes, selection is not expected to have any impact. Neutral markers will only be submitted to genetic drift and mutation, and the sampling effect due to the carrying capacity (Dobzhansky and Pavlovsky, 1957) is expected to be the main source of variability of such genetic diversity indicators. Mutation introduces new genetic diversity, but we do not expect it to be important in the analysis because we run the simulations over too few generations for the mutations to accumulate (Halligan and Keightley, 2009). At equilibrium in a finite size population only submitted to mutation, diversity will be maintained if $4 \times Ne \times \mu > 1$. Taking the carrying capacity as a rough value for the genetic effective size, only one set of parameter values (*carr_cap* = 1000 and *mutRate* = 10^{-3}) would meet the condition. Moreover, as the simulations are far from reaching equilibrium, the time period studied being very short compared to $4Ne$, mutation is not expected to contribute to variability in the outputs of the simulations (Crow, 1950).

2.3.2.2. Hypotheses for isolated populations, selected markers.

We analyse here the genetic diversity indicators at the selected markers. The evolutionary forces on the selected markers are the genetic drift, mutation and selection (Kirk and Freeland, 2011). The selection should affect the most the variability of the indicators as populations are generally initialised far from the populations optima. Genetic drift is also expected to affect the variability of the indicators due to cases with small population sizes. As for neutral markers, mutation is not expected to contribute much to variability.

2.3.2.3. Hypotheses for colonisation–extinction.

In these SAs, we analyse the genetic diversity at neutral and selected markers separately in cases where populations are connected by colonisation and submitted to extinction. As we add colonisation and extinction to the Isolated populations SAs, we expect the same genetic forces to contribute to the variability of the model, i.e. genetic drift and selection, together with colonisation and extinction. The importance of the contribution of each parameter is difficult to foresee due to the multiple parameters of the system, and little information is available from analytical approaches. Moreover, the demographic forces of the simulations and the network topologies are expected to contribute to most of the variability of the survival rate (Barbillon et al., 2015).

2.3.2.4. Hypotheses for migration.

In these SAs, the genetic diversity is analysed separately at neutral and selected markers in cases where populations are connected by migration. For genetic diversity parameters, the expectations are quite similar to those of Colonisation–Extinction SAs, as migration and colonisation both correspond to seed circulation among populations, therefore influencing both within and between-population genetic diversity indices. So, the same genetic forces are expected to contribute to the variability of the model, i.e. genetic drift and selection, as well as migration and network topologies, but without presuming the relative importance of the different parameters.

3. Results

3.1. Isolated populations

3.1.1. Isolated populations, neutral markers

The mean expected heterozygosity (Hs) in the SA Isolated populations, Neutral markers decreases from 0.5 to 0.4 over the thirty generations (Fig. 2(a)). Its variability also increases with 90% of the simulations between 0.5 and 0.25 Hs at the thirtieth generation.

The mean population differentiation (Gst) in the SA Isolated populations, Neutral markers increases from 0 to 0.25 (*cf.* Fig. 2(b)). The variability also increases during the simulations, with 90% of the replicates between 0 and 0.5.

The carrying capacity (*carr_cap* in the figures), i.e. the size of the populations, explains most of the variability of the expected heterozygosity indicator for the SA Isolated populations, Neutral markers (Fig. 2(a)). The remaining variability of the model is mostly explained by the reproduction regime (*percentSelf* in the figures). The same phenomenon can be observed for the population differentiation indicator (Gst) (Fig. 2(b)).

The same distribution of the variability is observed for both indicators in the last generation graph for the SA Isolated populations, Neutral markers, as shown in graph Fig. 2(c).

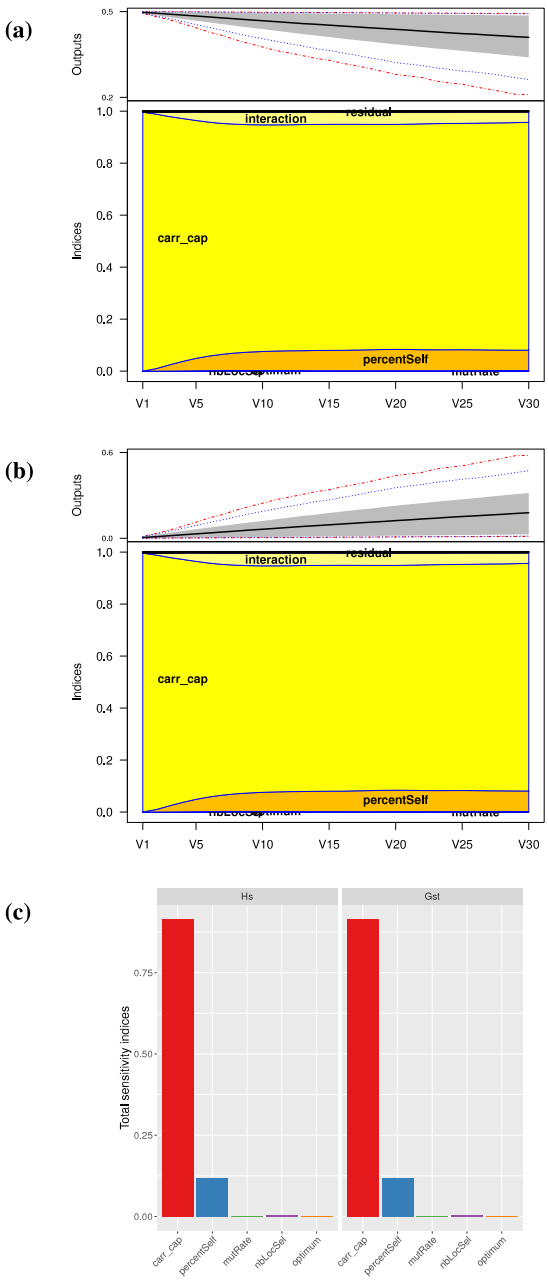

Fig. 2. Share of variability of the (**a**) Hs and (**b**) Gst along the generations of the simulations, and (**c**) Hs and Gst for the last generation of the simulations, induced by each parameter, for the SAs Isolated populations, Neutral markers. See details in Section 2.2.6 for interpretation.

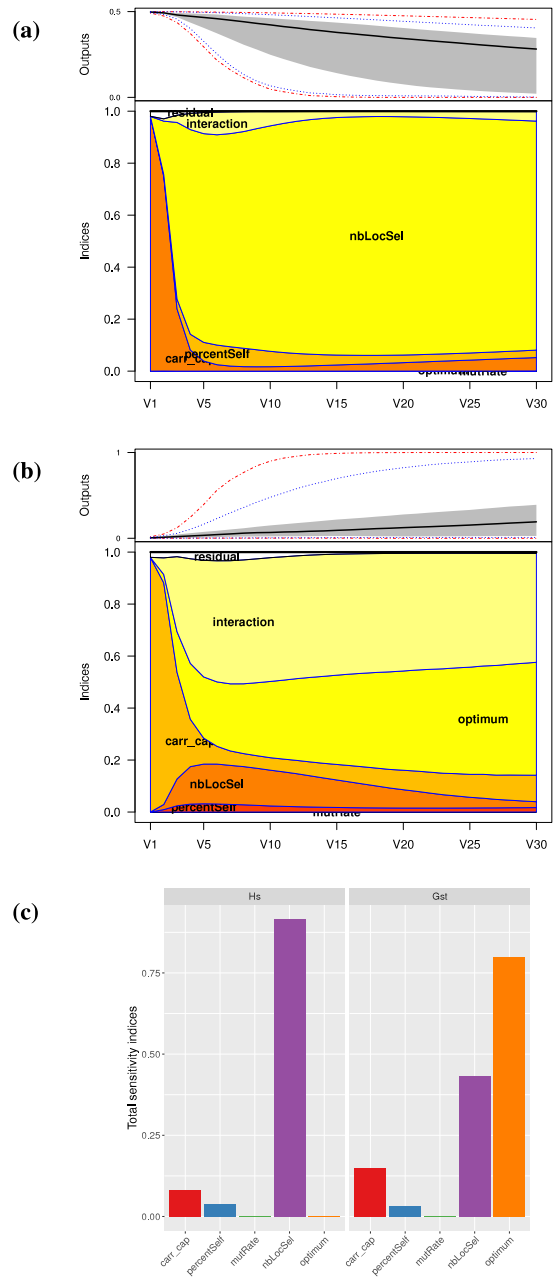

Fig. 3. Share of the variability of the (**a**) Hs and (**b**) Gst along the generations of the simulations, and (**c**) Hs and Gst for the last generation of the simulations, induced by each parameter, for the SAs Isolated populations, Selected markers. See details in Section 2.2.6 for interpretation.

3.1.2. Isolated populations, selected markers

The mean Hs indicator decreases from 0.5 to 0.3 over time (Fig. 3(a)) and its variability maximises at the fifteenth generation, with 90% of the replicates between slightly less than 0.5 and 0 at the last generation.

The mean Gst increases from 0 to 0.2 over time. Its variability maximises at generation 15, with 90% of the replicates between 0 and 0.95 at the last generation (Fig. 3(b)).

The number of loci under selection (*nbLocSel* in the figures) explains most of the variability of Hs (Fig. 3(a)).

During the first three generations, the carrying capacity explains transiently most of the variability of the indicator, later replaced by the reproduction regime.

Interactions explain almost half of the variability of the Gst indicator in Fig. 3(b), therefore the last generation graph (Fig. 3(c)) is important to determine the components of the variability. The optimum parameter (*optimum* in the figures) explains most of the variability of the Gst indicator. The number of loci under selection is the second most important contributor to the variability of this indicator, by mostly bringing it through interaction with other parameters. Finally, the carrying capacity and the reproduction regime both contribute to the variability of the Gst, especially in the early generations.

3.2. Colonisation–extinction

3.2.1. Colonisation–extinction, neutral markers

The mean theoretical heterozygosity for the SA Colonisation– Extinction, Neutral markers (Fig. 4(a)) decreases from 0.5 to 0.4 during the thirty generations of the simulations. 90% of Hs values range between

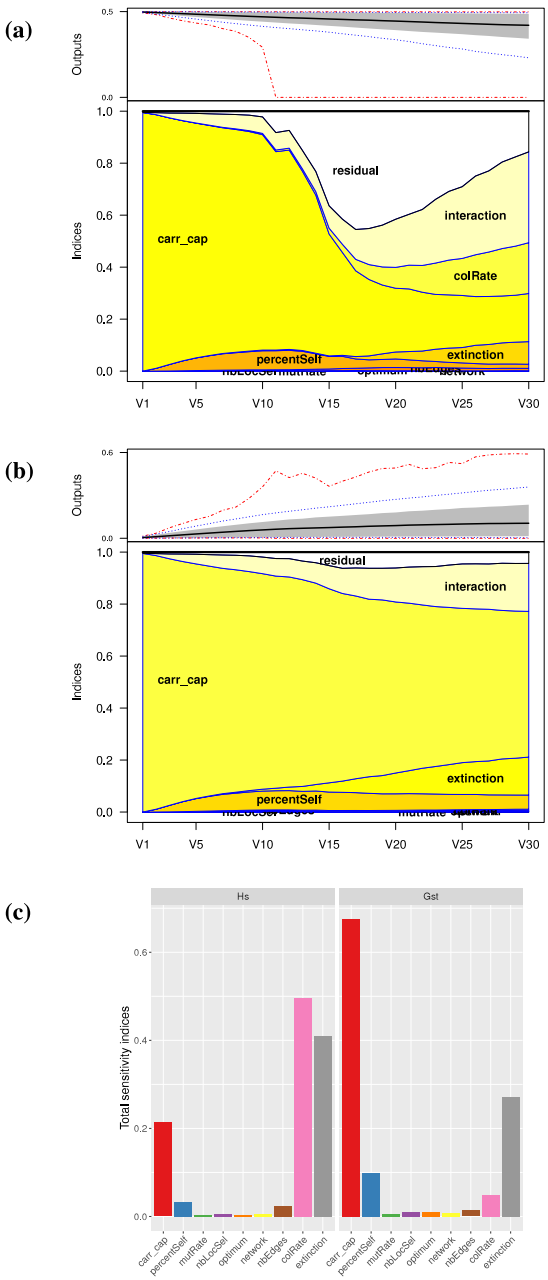

Fig. 4. Share of variability of the (**a**) Hs and (**b**) Gst along the generations of the simulations, and (**c**) Hs and Gst for the last generation of the simulations, induced by each parameter, for the SAs Colonisation–Extinction, Neutral markers. See details in Section 2.2.6 for interpretation.

0.5 and 0.25. Extreme values in the simulations appear at the thirteenth generation, corresponding to extinct or fixed metapopulations. The non continuity of the variability increases the share of residual up to 0.45 of the variability. When the variability becomes continuous again, the residual reduces to 0.2 of the variability share.

The carrying capacity explains the largest part of the variability of Hs in the first fifteen generations, while the reproduction regime explains the remaining part. The Figs. 2(a) and 4(a) display similar trends for the variability share during the first fifteen generations.

From the fifteenth generation to the thirtieth, the colonisation rate (*colRate* in the figures) and the extinction rate (*extinction* in the figures) take a larger part in the variability of the Hs indicator. It is likely that particular combinations of extinction and colonisation rates (high extinction and low colonisation) lead to extinct metapopulations with $Hs = 0$, inducing a large variability in the outputs. These two parameters are the main components of the interaction, as it can be seen in the Fig. 4(c).

The mean population differentiation for the SA Colonisation– Extinction, Neutral markers (Fig. 4(b)) increases from 0 to 0.1 during the thirty generations. 90% of the Gst range between 0 and 0.4.

Similarly to the trends observed in the Fig. 2(b), the carrying capacity explains most of the variability of the Gst indicator, especially in the first twenty generations. The reproduction regime is the second most important parameter for the same generations. During the last ten generations, the share of the carrying capacity decreases, and the extinction parameter becomes the second most important parameter, as shown in Fig. 4(c).

3.2.2. Colonisation–extinction, selected markers

The mean expected heterozygosity for the SA Colonisation–Extinction, Selected markers (Fig. 5(a)) decreases from 0.5 to 0.3 during the thirty generations, i.e. as expected much more than at neutral markers. 90% of Hs range between 0.5 and 0.

The number of locus under selection explains the largest part of the variability of the expected heterozygosity, especially from the third generation on. The reproduction regime is the second most important parameter from the generation four to twenty, then replaced by the colonisation rate. The extinction rate is also important in the last generations, as shown in Fig. 5(c), though it is mostly influencing through interactions with other parameters.

The mean population differentiation increases from 0 to 0.1 through the thirty generations. 90% of the simulations range from 0 to 1.

Most of the variability of the Gst indicator (Fig. 5(b)) is explained by the interaction between parameters after the fifth generation. In the last generation, the interaction is mainly composed by the number of locus under selection, and in a second order by the colonisation rate and the extinction rate, as it can be seen in Fig. 5(c). The following most important parameters are the optima of the populations and the carrying capacity.

3.2.3. Colonisation–extinction demographic indices

The survival rate decreases from 1 to 0.8 through the thirty generations (Fig. 6). Moreover, the variability of the indicator rapidly increases, and 90% of the simulations range between 1 and 0.04.

The colonisation rate explains less than half of the variability of the indicator during the first half of the generations. Its share increases over the thirty generations, and ends explaining the majority of the variability of the survival rate, including in interaction with other parameters.

The extinction rate explains half of the variability of the survival rate in the early generations. Its share decreases during the thirty generations to stand below the colonisation rate.

The number of edges (*nbEdges* in the figures), i.e. the number of social links through which seeds can circulate, shows a steady share of 0.05 of the variability of the survival rate.

3.3. Migration

The mean expected heterozygosity for the SA Migration, Neutral markers (Fig. A.1) decreases from 0.5 to 0.45 during the thirty generations. The population differentiation of the same SA increases from 0 to 0.05.

For the SA Migration, Neutral markers, the carrying capacity explains most of the variability of Hs and Gst indicators (Fig. 7(a)). The following most important parameters are the migration rate (*migRate* in the figure) and the proportion of seeds replaced by migration (*migReplace* in the figure). The reproduction regime explains 0.1 of the variability. Finally, the number of edges brings half the variability of the reproduction regime to the simulation. Compared to the Colonisation–Extinction,

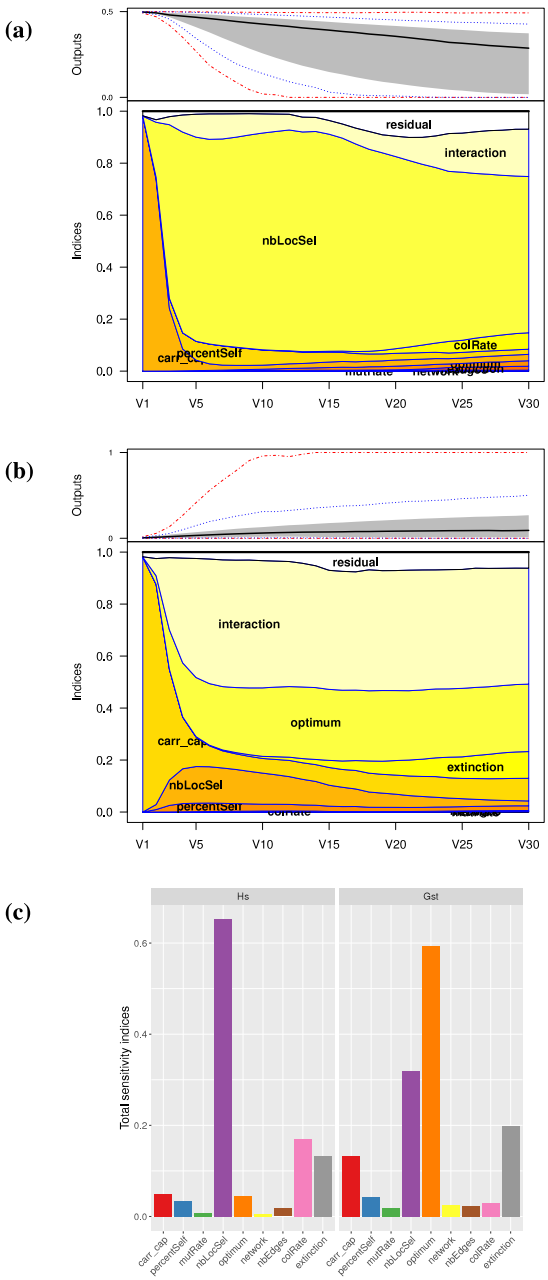

Fig. 5. Share of variability of the (**a**) Hs and (**b**) Gst along the generations of the simulations, and (**c**) Hs and Gst for the last generation of the simulations, induced by each parameter, for the SAs Colonisation–Extinction, Selected markers. See details in Section 2.2.6 for interpretation.

Neutral markers SA, seed circulation parameters induce less variability in the output as they do not lead to extinct metapopulations.

The mean expected heterozygosity for the SA Migration, Selected markers (Fig. A.2) decreases from 0.5 to 0.35 during the thirty generations. The population differentiation of the same SA increases from 0 to 0.05.

For the SA Migration, Selected markers, the number of locus submitted to selection explains the largest part of the variability of the Hs indicator at the last generation (Fig. 7(b)). The distribution of optima represents a source of variability over 7 fold lower. The migration rate, the proportion of seed replaced during a migration, the carrying capacity and the reproduction regime also bring variability to the simulations to a lesser degree.

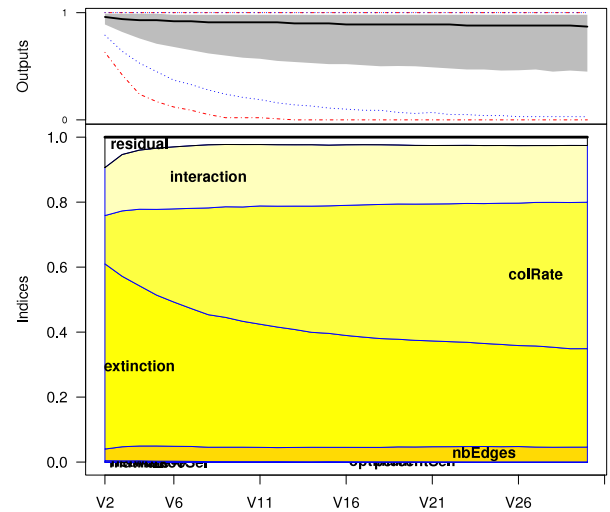

Fig. 6. Share of variability of the survival rate induced by each parameter along the generations of the simulations for the SAs Colonisation–Extinction. See details in Section 2.2.6 for interpretation.

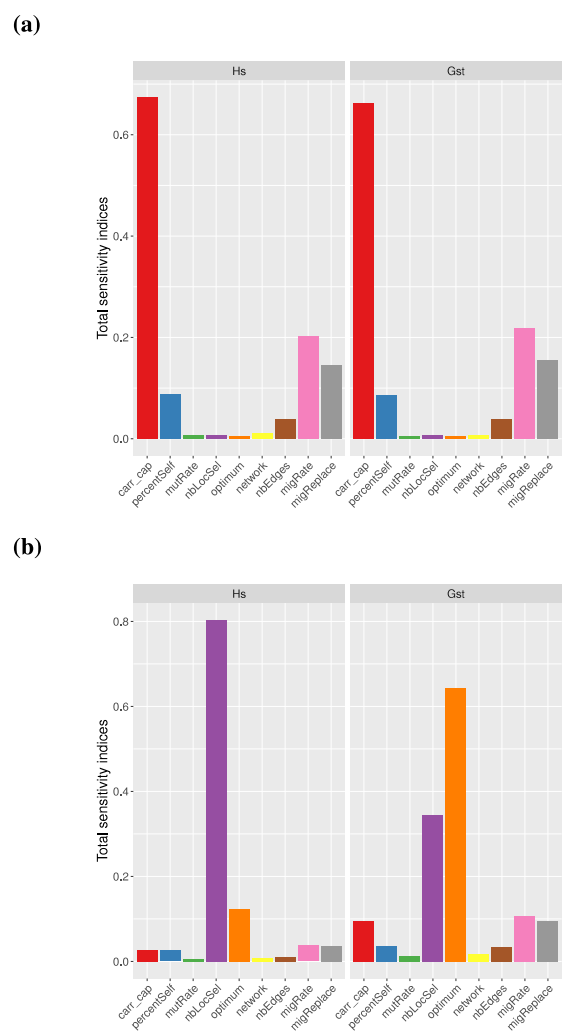

Fig. 7. Share of variability induced by each parameter of the last generation for the (**a**) Neutral markers and (**b**) Selected markers for the SAs Migration. See details in Section 2.2.6 for interpretation.

For the Migration, Selected markers SA, the distribution of selective optima of the populations and, in a lower extent, the number of locus under selection, explain the greatest part of the variability of the Gst indicator (Fig. 7(b)). The migration rate, the proportion of seeds replaced by migration, and the carrying capacity, explain ten times less of the variability of Gst than the optimum distribution. Finally, most of the remaining variability is explained by the reproduction regime, the number of edges of the network, and the network topology (*network* in the figures).

4. Discussion

4.1. Influence of the parameters in the different SAs

4.1.1. Origin of the variability for isolated populations

Neutral markers of isolated populations (*cf* Section 3.1.1 for the description of the SAs) are mainly submitted to genetic drift parameters. The carrying capacity explains most of the variability of the outputs of the model, and the reproduction regime is the second most important parameter. The mutation rate (*mutRate* in the figures), does not induce variability in the outputs of the model, which is consistent with the expectations (Halligan and Keightley, 2009; Schoen and Brown, 1991), as genetic drift is the main evolutionary force expected on neutral markers given the rather small population sizes considered and the short time period studied (Dobzhansky and Pavlovsky, 1957). The selection parameters do not bring any variability in the outputs which is consistent with the fact that all loci markers are neutral and modelled as unlinked genetically to the selected markers and so no linkage drag is expected to occur. Although populations are initialised by drawing alleles at random independently at each locus, sampling of a small number of multilocus genotypes in the case of small population size (40) could generate some initial linkage disequilibrium (Ohta and Kimura, 1969). This could generate variability due to the interaction between carrying capacity and selection parameters. However, this does not seem to be the case here.

In contrast, Selected markers (*cf* Section 3.1.2 for the description of the SA) are mostly under the influence of the selection forces. The number of locus under selection explains most of the variability of the Hs indicator. This is consistent with our hypotheses (see 2.3.2.2) and is caused by the fact that the response to selection will be faster when fewer loci determine the fitness, as little linkage disequilibrium is expected among positive and negative alleles at different selected markers (Kirk and Freeland, 2011). Although selection is expected to be stronger in larger populations (i.e. when genetic drift is limited) than in small populations, no interaction seems to influence the variability of Hs indicator. For the Gst indicator, the main parameter that explains the variability of the simulations is the distribution of the selection optima of the populations. This is consistent with our expectations (see 2.3.2.2) as the distribution of the optima will directly impact the differentiation among populations. This is confirmed by the fact that the optimum parameter has strong interactions with the number of locus under selection.

For both types of markers (neutral and selected), Hs decreases over time as expected, especially as genetic drift is strong and as selection is intense for the selected markers. In parallel, Gst increases at both types of markers. The increase in differentiation is larger at neutral markers as expected (Wang, 2015) because the sampling effect will not pick the same alleles in the different populations. At selected markers, the differentiation among populations depends on the optimum distribution and number of locus under selection. While similar optima for all populations will lead to few or no differentiation, contrasted and continuous optima can maximise differentiation depending on the number of selected locus.

In the case of isolated populations, the genetic drift parameter (i.e. the carrying capacity of the populations) and the two selection parameters (number of selected loci and distribution of the populations optima) will the most influence the variability of the genetic diversity outputs depending on the type of markers we consider (neutral or selected), showing therefore the importance to tune these parameters finely.

4.1.2. Origin of the variability for colonisation–extinction

The genetic diversity indicators in the Colonisation–Extinction SAs are mainly influenced by the genetic drift or selection parameters depending on the markers analysed.

The colonisation and extinction rates are of second importance to induce variability in the outputs of the model in particular at neutral markers. The colonisation rate mostly influences the within-population genetic diversity through its capacity to introduce locally new diversity due to the possibility of multiple sources of seed flows, while mainly the extinction rate influences the genetic differentiation among populations.

It is surprising that the network parameters contribute so little to the variability of the outputs of the model. This could be explained by the large range of variation of the genetic drift parameter, which accommodates most of the variability. In contrast, the colonisation and extinction parameters explain all the variability of the survival rate. This is expected (see 2.3.2.3) as these parameters and their relative values have a direct impact on the demography of the metapopulation.

In these SAs, the evolution of Hs mean is quite similar to the evolution in Isolated Populations SAs whatever the markers, while the genetic differentiation increases much less over time due to seed circulation among populations. Thus, even though the existence of seed flows between populations influences the evolution over generations of the genetic differentiation between populations, neither the colonisation rate nor the network specificities strongly determine the variability of Gst response.

4.1.3. Origin of the variability for migration

The variability of genetic diversity indicators in the Migration SAs is mainly influenced by the genetic drift or selection parameters depending on the markers analysed, as is the case for the Isolated populations SAs. The migration parameters are of second order of importance to explain the variability of the output of the model.

The migration rate induces more variability in the outputs of the model than the colonisation rate in Colonisation–Extinction SAs for the Gst of Neutral markers. Yet, as seed circulation under extinction–colonisation regime is governed by the combination of both the extinction rate and the colonisation rate (as colonisation can only occur in case a population became extinct), one should consider the influence of these two parameters together. Similarly, it is sensible to consider the influence of both the migration rate and the replacement rate together as, under the migration regime, seed circulation is governed by the combination of the migration rate and the replacement rate within the target population.

In the same way as the extinction and colonisation parameters in the Colonisation–Extinction SAs, the migration rate and the replacement rate parameters contribute both equally to the variability of the Selected markers indicators but to a smaller extent compared to variability of the Neutral markers indicators.

Finally, network parameters have very few impacts on the variability of the genetic outputs. As for the Colonisation–Extinction SAs, this might be explained by the wide range of variation of our genetic drift parameters.

In the Migration SAs, Hs mean over generations decreases slightly less compared to the trends in Isolated population and Colonisation–Extinction SAs while the mean Gst increases less than in Colonisation–Extinction SA indicating that the range of situations considered in Migration SAs leads to less structure of the genetic diversity within the metapopulation.

Consistently, seed flows between populations described here by the migration rate and the replacement rate do not strongly determine the variability of genetic diversity and population differentiation.

4.1.4. Early generations of colonisation–extinction, neutral markers

The residual of Hs increases very fast after the tenth generation, and then reduces progressively (Fig. 4(a)). This is due to the fact that for certain combinations of parameters, at this generation, in all the replicates the metapopulation is empty or fixed at all loci. This causes a huge variability in the outputs that the analysis cannot explain by any up to second order effect or interaction. Afterwards other combinations of parameters lead to an empty or fixed metapopulation, and the variability turns more continuous, so that the analysis becomes more explanatory, and thus the residual decreases.

It can be noted that the colonisation rate and the extinction rate have a larger share in the variability of the outputs in the late generations than in the early generations. This is because in the early generations, the populations are not genetically stabilised yet to their reproduction regime and population size. The genetic frequencies will thus vary a lot in the early generations, and only stabilise (usually) around the 15th generation. The "stabilised share" of the parameters to the variability of the outputs can then be observed.

It is interesting to note that colonisation does not simply behave as migration. The migration rate seems to influence more the variability of the Gst indicator than the colonisation rate does (Figs. 7, 4(c) and 5(c)). This might come from the fact that the migration is not dependent of an extinction process to move seeds. Migration could thus take place more often than colonisation, creating more gene flow between populations than colonisation.

It is surprising that so little variability of the outputs in the Colonisation–Extinction and in the Migration SAs is coming from differences in the topology of the network, compared to the variability due to the genetic drift parameter and to extinction, colonisation and migration rates.

In Barbillon et al. (2015), the authors evaluated with a sensitivity analysis a dynamic colonisation extinction model in the context of metapopulations to assess the impact of several parameters related to colonisation, extinction, network topology and density on the average survival rate of a variety over all the farms growing the variety (crop metapopulation) and on the persistence of the variety in the crop metapopulation. The authors show that after the extinction and colonisation parameters, it is the number of edges in the network that induces most of the variability. Then, the network topology induces less but not negligible variability. We show here that the same parameters induce far less variability. The added genetic aspect or the large variation range of the genetic drift parameters might explain this observed difference (Tan et al., 2017).

4.2. Limitations of the experimental design

Even if we used a fractional factorial design of experiments, increasing the number of parameters beyond 9 would result in too many simulations for the duration of the simulations (which ranges from a few dozens of seconds to more than 24 h). Moreover, running six different SAs allowed us to better understand the dynamics of the different evolutionary forces separately. We thus chose to restrict the SAs dedicated to Isolated populations to 5 parameters to study evolutionary forces related to genetic drift and selection, and the SAs dedicated to Colonisation–Extinction and Migration to 9 parameters to study forces related to colonisation and migration respectively, along with the genetic drift and selection.

As mentioned in the Section 2.2.1, using "planor" requires the same number of levels for each parameter chosen. As we chose to have 3 levels for our parameters (i.e. two extreme values and an intermediate one), it imposed us to chose parameters among the continuous ones, and leave aside the "on/off" parameters even if they might have been interesting to include in the SA.

Using a fractional factorial design of experiments allows the number of simulations to be reduced considerably, in our case up to 10 times less. However, it does not allow to detect interactions above the second order. This limits the power of the analysis, although we observe in the results that very little variability can be explained by second order interactions. Already, few of the variability is explained by second order interactions. This indicates that while we do not yet fully understand the role of each parameter in the variability of the indicators, we expect very little divergence between what we understand of the model and what actually happens in it. In addition, interactions above second order, not detected by the analysis, are included in the residual. Thus, as long as the residual is small, one can be sure that the larger interactions induce very little variability in the model outputs.

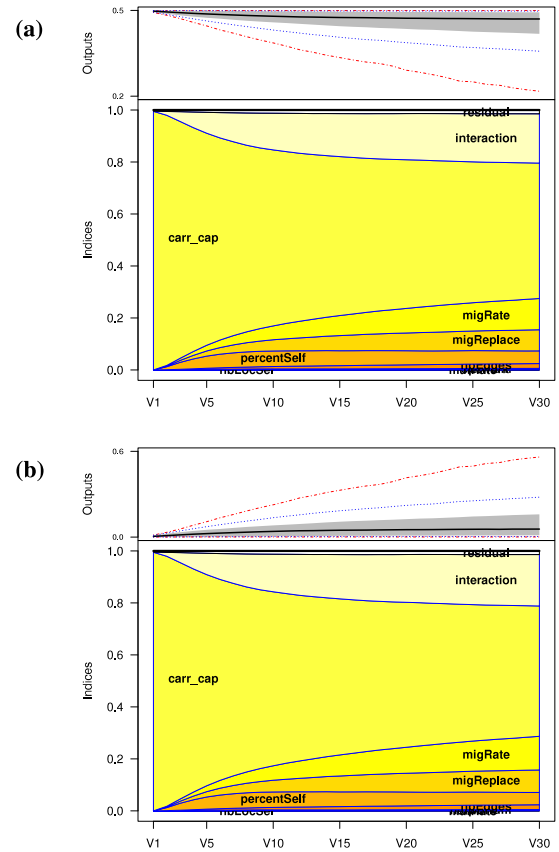

Fig. A.1. Share of variability of the (**a**) Hs and (**b**) Gst along the generations of the simulations induced by each parameter, for the SAs Migration, Neutral markers. See details in Section 2.2.6 for interpretation.

The use of two separate SAs for selected and non-selected markers allowed a better understanding of the behaviour of the model with and without selection, as when the number of selected loci is set to 0 (i.e. no selection), it is not possible to define selective optima for populations or to analyse genetic diversity at selected markers. Yet, we are not able to quantitatively compare the results of the two SAs, because the combinations of parameters might interact differently in the two SAs. We are, however, able to qualitatively compare two SAs with each other. In order to diversify as much as possible the networks used by CropMetaPop in the analysis, we generated a library of a hundred networks for each combination of network type and density. Therefore, as we run ten replicates, it is quite unlikely to find the same network several times in the same parameter combination.

To analyse the results of the simulation more efficiently, home-made Python scripts were created in the team. It allows for an adapted and fast computation of the genetic and demographic indicators.

5. Conclusion

Here we present CropMetaPop, a model that addresses the lack of a suitable model to study the genetic evolution of crop metapopulations. Critical features of crop metapopulations that can usually

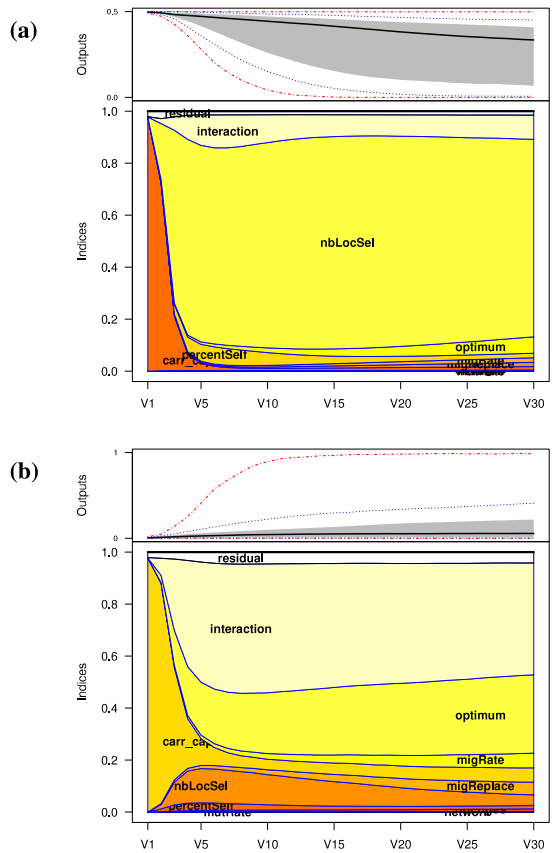

Fig. A.2. Share of variability of the (**a**) Hs and (**b**) Gst along the generations of the simulations induced by each parameter, for the SAs Migration, Selected markers. See details in Section 2.2.6 for interpretation.

not be represented in classic metapopulations models, are included in CropMetaPop. Among those features, the distinct migration and colonisation processes, the fact that these processes happen on seeds and not on juveniles and/or adults. Moreover the CropMetaPop allows to model precisely the seed movements that can be realised. This seed circulation network can be defined precisely based on real data or defined randomly. Moreover, CropMetaPop allows to set parameters specific to seed movements, such as the ratio of seeds replaced by migration, or the ratio kept by farmers haring seeds, among other parameters.

We performed sensitivity analyses of the CropMetaPop model for a time frame corresponding to the situations encountered in the field, i.e. a few dozen generations maximum, to check that the model works as expected, and to determine the relative influence that the major input parameters have on the outputs of the model. Results showed that the drift-related parameter was of the most influential on the variability of the genetic diversity indicators, especially at the neutral markers, while indicators at selected markers were mostly influenced by the number of loci under selection and the distribution of the populations optimums. Colonisation–extinction and migration processes through the rates of extinction, colonisation, migration and migration replacement, introduced additional variability to the outputs of the model in the corresponding sensitivity analyses, while the topology of the network parameters induced only little variability in the outputs of the model. The results of the sensitivity analyses are not intended to directly describe situations encountered in the field. Rather, they will Rather, they will help CropMetaPop users to be vigilant in the choice of values for certain parameters, especially those that play an important role in the results. These results support a proper functioning of the model. CropMetaPop can therefore be used in concrete applications to study the genetic evolution of crop metapopulations. For example, the model can be used to practices and organisational modes on the evolution of crop genetic diversity. This type of work could be used as a mediation tool to build decisions in a collective managing crop genetic diversity.

CRediT authorship contribution statement

Baptiste Rouger: Improved the code of the software, Designed, ran and analysed the sensitivity analyses, Writing – original draft. **Isabelle Goldringer:** Conceptualised and designed the software, Designed the sensitivity analyses and discussed the results, Contributed to and edited the manuscript. **Pierre Barbillon:** Designed the sensitivity analyses, Edited the manuscript. **Anne Miramon:** Developped the code of the software. **Abdel Kader Naino Jika:** Contributed to test the model, Contributed to part of the manuscript. **Mathieu Thomas:** Conceptualised and designed the software, Designed the sensitivity analyses and discussed the results, Contributed to and edited the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank Frédéric Hospital for his help in designing and developing the software. The authors are very grateful to the MIRES network financially supported by INRAE and to François Massol for his very helpful advises. This research was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 633571 (DIVERSIFOOD project) for the period 2015–2019 and BR received a grant from the École Doctorale Frontières de l'Innovation en Recherche et Éducation - Programme Bettencourt and PhD funding from INRAE.

Appendix. Results of SAs migration

See Figs. A.1 and A.2.

References

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512. http://dx.doi.org/10.1126/science.286.5439.509, URL: https://www.sciencemag.org/lookup/doi/10.1126/science.286.5439.509.

Barbillon, P., Thomas, M., Goldringer, I., Hospital, F., Robin, S., 2015. Network impact on persistence in a finite population dynamic diffusion model: Application to an emergent seed exchange network. J. Theoret. Biol. 365, 365–376. http://dx.doi.org/10.1016/j.jtbi.2014.10.032, URL: http://www.sciencedirect.com/science/article/pii/S0022519314006225.

Bellon, M.R., Hodson, D., Hellin, J., 2011. Assessing the vulnerability of traditional maize seed systems in Mexico to climate change. Proc. Natl. Acad. Sci. 108 (33), 13432–13437. http://dx.doi.org/10.1073/pnas.1103373108, URL: http://www.pnas.org/content/108/33/13432.abstract.

Bidot, C., Lamboni, M., Monod, H., 2018. Multisensi: multivariate sensitivity analysis. R package version 2.1-1. URL: https://CRAN.R-project.org/package=multisensi.

Boster, J.S., 1986. Exchange of varieties and information between Aguaruna manioc cultivators. Am. Anthropol. 88 (2), 428–436.

Burgers, S.L.G.E., Hofstede, G.J., Jonker, C.M., Verwaart, T., 2010. Sensitivity analysis of an agent-based model of culture's consequences for trade. In: Li Calzi, M., Milone, L., Pellizzari, P. (Eds.), Progress in Artificial Economics. Vol. 645. Springer, Berlin, Heidelberg, pp. 253–264. http://dx.doi.org/10.1007/978-3-642-13947-5_21, URL: http://link.springer.com/10.1007/978-3-642-13947-5_21.

Calvet-Mir, L., Calvet-Mir, M., Molina, J.L., Reyes-García, V., 2012. Seed exchange as an agrobiodiversity conservation mechanism. a case study in vall Fosca, Catalan Pyrenees, Iberian Peninsula. Ecol. Soc. 17 (1), http://dx.doi.org/10.5751/ES-04682-170129, URL: http://www.ecologyandsociety.org/vol17/iss1/art29/.

Cariboni, J., Gatelli, D., Liska, R., Saltelli, A., 2007. The role of sensitivity analysis in ecological modelling. Ecol. Model. 203 (1–2), 167–182. http://dx.doi.org/10.1016/j.ecolmodel.2005.10.045, URL: https://linkinghub.elsevier.com/retrieve/pii/S0304380006005734.

Carpani, M., Bergez, J.-E., Monod, H., 2012. Sensitivity analysis of a hierarchical qualitative model for sustainability assessment of cropping systems. Environ. Model. Softw. 27–28, 15–22. http://dx.doi.org/10.1016/j.envsoft.2011.10.002, URL: https://linkinghub.elsevier.com/retrieve/pii/S1364815211002131.

Crow, J.F., 1950. An introduction to population genetics. Am. J. Hum. Genet. 2 (3), 279–280, URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1716346/.

Dobzhansky, T., Pavlovsky, O., 1957. An experimental study of interaction between genetic drift and natural selection. Evolution 11 (3), 311. http://dx.doi.org/10.2307/2405795, URL: https://www.jstor.org/stable/2405795?origin=crossref.

Erdős, P., Rényi, A., 1959. On random graphs. Publ. Math. 6 (26), 290–297.

Fenzi, M., Rogé, P., Cruz-Estrada, A., Tuxill, J., Jarvis, D., 2021. Community seed network in an era of climate change: dynamics of maize diversity in yucatán, Mexico. Agric. Hum. Values http://dx.doi.org/10.1007/s10460-021-10249-3.

Fuentes, F.F., Bazile, D., Bhargava, A., Martínez, E.A., 2012. Implications of farmers' seed exchanges for on-farm conservation of quinoa, as revealed by its genetic diversity in Chile. J. Agric. Sci. 150 (6), 702–716. http://dx.doi.org/10.1017/S0021859612000056, URL: https://www.cambridge.org/core/journals/journal-of-agricultural-science/article/implications-of-farmers-seed-exchanges-for-onfarm-conservation-of-quinoa-as-revealed-by-its-genetic-diversity-in-chile/E8DF48B0E561A371ED4DFB1DB1C422AD.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. 99 (12), 7821–7826. http://dx.doi.org/10.1073/pnas.122653799, URL: http://www.pnas.org/cgi/doi/10.1073/pnas.122653799.

Halligan, D.L., Keightley, P.D., 2009. Spontaneous mutation accumulation studies in evolutionary genetics. Annu. Rev. Ecol. Evol. Syst. 40 (1), 151–172. http://dx.doi.org/10.1146/annurev.ecolsys.39.110707.173437, URL: https://www.annualreviews.org/doi/10.1146/annurev.ecolsys.39.110707.173437.

van Heerwaarden, J., van Eeuwijk, F.A., Ross-Ibarra, J., 2010. Genetic diversity in a crop metapopulation. Heredity 104 (1), 28–39. http://dx.doi.org/10.1038/hdy.2009.110, URL: https://www.nature.com/articles/hdy2009110.

Johnson, T., Barton, N., 2005. Theoretical models of selection and mutation on quantitative traits. Philos. Trans. R. Soc. B 360 (1459), 1411–1425. http://dx.doi.org/10.1098/rstb.2005.1667, URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2005.1667.

Kimura, M., 1980. Average time until fixation of a mutant allele in a finite population under continued mutation pressure: Studies by analytical, numerical, and pseudo-sampling methods. Proc. Natl. Acad. Sci. 77 (1), 522–526. http://dx.doi.org/10.1073/pnas.77.1.522, URL: http://www.pnas.org/cgi/doi/10.1073/pnas.77.1.522.

Kirk, H., Freeland, J.R., 2011. Applications and implications of neutral versus non-neutral markers in molecular ecology. Int. J. Mol. Sci. 12 (6), 3966–3988. http://dx.doi.org/10.3390/ijms12063966, URL: http://www.mdpi.com/1422-0067/12/6/3966.

Kobilinsky, A., 2005. PLANOR: program for automatic generation of regular experimental designs.

Kobilinsky, A., Bouvier, A., Monod, H., 2019. PLANOR: an R package for the automatic generation of regular fractional factorial designs.

Kobilinsky, A., Monod, H., Bailey, R., 2017. Automatic generation of generalised regular factorial designs. Comput. Statist. Data Anal. 113, 311–329. http://dx.doi.org/10.1016/j.csda.2016.09.003, URL: https://linkinghub.elsevier.com/retrieve/pii/S0167947316302110.

Labeyrie, V., Thomas, M., Muthamia, Z.K., Leclerc, C., 2016. Seed exchange networks, ethnicity, and sorghum diversity. Proc. Natl. Acad. Sci. 113 (1), 98–103. http://dx.doi.org/10.1073/pnas.1513238112, URL: https://www.pnas.org/content/113/1/98.

Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., Monod, H., 2009. Multivariate global sensitivity analysis for dynamic crop models. Field Crops Res. 113 (3), 312–320. http://dx.doi.org/10.1016/j.fcr.2009.06.007, URL: http://www.sciencedirect.com/science/article/pii/S0378429009001531.

Lamboni, M., Monod, H., Makowski, D., 2011. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. Reliab. Eng. Syst. Saf. 96 (4), 450–459. http://dx.doi.org/10.1016/j.ress.2010.12.002, URL: http://www.sciencedirect.com/science/article/pii/S0951832010002504.

Louette, D., Charrier, A., Berthaud, J., 1997. In situ conservation of maize in Mexico: Genetic diversity and maize seed management in a traditional community. Econ. Botany 51 (1), 20–38. http://dx.doi.org/10.1007/BF02910401, WOS:A1997WR70800003.

Lurette, A., Touzeau, S., Lamboni, M., Monod, H., 2009. Sensitivity analysis to identify key parameters influencing Salmonella infection dynamics in a pig batch. J. Theoret. Biol. 258 (1), 43–52. http://dx.doi.org/10.1016/j.jtbi.2009.01.026, URL: https://linkinghub.elsevier.com/retrieve/pii/S0022519309000307.

Massol, F., Débarre, F., 2015. Evolution of dispersal in spatially and temporally variable environments: The importance of life cycles: dispersal evolution and life cycles. Evolution 69 (7), 1925–1937. http://dx.doi.org/10.1111/evo.12699, URL: https://onlinelibrary.wiley.com/doi/10.1111/evo.12699.

McGuire, S.J., 2008. Securing access to seed: Social relations and sorghum seed exchange in eastern ethiopia. Hum. Ecol. 36 (2), 217–229. http://dx.doi.org/10.1007/s10745-007-9143-4.

Naino Jika, A.K., Dussert, Y., Raimond, C., Garine, E., Luxereau, A., Takvorian, N., Djermakoye, R.S., Adam, T., Robert, T., 2017. Unexpected pattern of pearl millet genetic diversity among ethno-linguistic groups in the Lake Chad Basin. Heredity http://dx.doi.org/10.1038/hdy.2016.128, URL: http://www.nature.com/hdy/journal/vaop/ncurrent/full/hdy2016128a.html.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nei, M., Maruyama, T., Chakraborty, R., 1975. The bottleneck effect and genetic variability in populations. Evolution 29 (1), 1. http://dx.doi.org/10.2307/2407137, URL: https://www.jstor.org/stable/2407137?origin=crossref.

Neuenschwander, S., Michaud, F., Goudet, J., 2019. QuantiNemo 2: a swiss knife to simulate complex demographic and genetic scenarios, forward and backward in time. Bioinformatics 35 (5), 886–888. http://dx.doi.org/10.1093/bioinformatics/bty737, Publisher: Oxford Academic. URL: https://academic.oup.com/bioinformatics/article/35/5/886/5078477.

Nowicki, K., Snijders, T.A.B., 2001. Estimation and prediction for stochastic block-structures. J. Amer. Statist. Assoc. 96 (455), 1077–1087. http://dx.doi.org/10.1198/016214501753208735, URL: http://www.tandfonline.com/doi/abs/10.1198/016214501753208735.

Ohta, T., Kimura, M., 1969. Linkage disequilibrium due to random genetic drift. Genet. Res. 13 (1), 47–55. http://dx.doi.org/10.1017/S001667230000272X, URL: https://www.cambridge.org/core/product/identifier/S001667230000272X/type/journal_article.

Pautasso, M., Aistara, G., Barnaud, A., Caillon, S., Clouvel, P., Coomes, O.T., Delêtre, M., Demeulenaere, E., De Santis, P., Döring, T., Eloy, L., Emperaire, L., Garine, E., Goldringer, I., Jarvis, D., Joly, H.I., Leclerc, C., Louafi, S., Martin, P., Massol, F., McGuire, S., McKey, D., Padoch, C., Soler, C., Thomas, M., Tramontini, S., 2013. Seed exchange networks for agrobiodiversity conservation. A review. Agron. Sustain. Dev. 33 (1), 151–175. http://dx.doi.org/10.1007/s13593-012-0089-6.

Peng, B., Amos, C.I., 2008. Forward-time simulations of non-random mating populations using simupop. Bioinformatics 24 (11), 1408–1409. http://dx.doi.org/10.1093/bioinformatics/btn179, URL: https://academic.oup.com/bioinformatics/article/24/11/1408/192516.

Peng, B., Kimmel, M., 2005. simuPOP : a forward-time population genetics simulation environment. Bioinformatics 21 (18), 3686–3687. http://dx.doi.org/10.1093/bioinformatics/bti584, URL: https://academic.oup.com/bioinformatics/article/21/18/3686/202296.

Pressoir, G., Berthaud, J., 2004. Patterns of population structure in maize landraces from the central Valleys of Oaxaca in Mexico. Heredity 92 (2), 88–94. http://dx.doi.org/10.1038/sj.hdy.6800387, URL: http://www.nature.com/articles/6800387.

Raquin, A.-L., Depaulis, F., Lambert, A., Galic, N., Brabant, P., Goldringer, I., 2008. Experimental estimation of mutation rates in a wheat population with a gene genealogy approach. Genetics 179 (4), 2195–2211. http://dx.doi.org/10.1534/genetics.107.071332, URL: http://www.genetics.org/cgi/content/abstract/179/4/2195.

Ravigne, V., Olivieri, I., Dieckmann, U., 2004. Implications of habitat choice for protected polymorphisms. Evol. Ecol. Res. 6 (1), 125–145, Number: 1 Publisher: Evolutionary Ecology, Ltd. URL: http://www.evolutionary-ecology.com/issues/v06n01/kkar1602.pdf.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environ. Model. Softw. 114, 29–39. http://dx.doi.org/10.1016/j.envsoft.2019.01.012, URL: http://www.sciencedirect.com/science/article/pii/S1364815218302822.

Schoen, D.J., Brown, A.H., 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proc. Natl. Acad. Sci. 88 (10), 4494–4497. http://dx.doi.org/10.1073/pnas.88.10.4494, URL: http://www.pnas.org/cgi/doi/10.1073/pnas.88.10.4494.

Schoen, D.J., Schultz, S.T., 2019. Somatic mutation and evolution in plants. Annu. Rev. Ecol. Evol. Syst. 50 (1), 49–73. http://dx.doi.org/10.1146/annurev-ecolsys-110218-024955, URL: https://www.annualreviews.org/doi/10.1146/annurev-ecolsys-110218-024955.

Song, Y., Fang, Q., Jarvis, D., Bai, K., Liu, D., Feng, J., Long, C., 2019. Network analysis of seed flow, a traditional method for conserving tartary buckwheat (fagopyrum tataricum) landraces in Liangshan, Southwest China. Sustainability 11 (16), 1–14, Publisher: MDPI. URL: https://ideas.repec.org/a/gam/jsusta/v11y2019i16p4263-d255458.html.

Takahata, N., Nei, M., 1984. F and g statistics in the finite island model. Genetics 107 (3), 501–504.

Tan, J., Cui, Y., Luo, Y., 2017. Assessment of uncertainty and sensitivity analyses for ORYZA model under different ranges of parameter variation. Eur. J. Agron. 91, 54–62. http://dx.doi.org/10.1016/j.eja.2017.09.001, URL: https://linkinghub.elsevier.com/retrieve/pii/S1161030117301235.

Thomas, M., Caillon, S., 2016. Effects of farmer social status and plant biocultural value on seed circulation networks in Vanuatu. Ecol. Soc. 21 (2), http://dx.doi.org/10.5751/ES-08378-210213.

Violon, C., Thomas, M., Garine, E., 2016. Good year, bad year: changing strategies, changing networks? A two-year study on seed acquisition in northern Cameroon. Ecol. Soc. 21 (2), http://dx.doi.org/10.5751/ES-08376-210234.

Wang, J., 2015. Does g_{ST} underestimate genetic differentiation from marker data? Mol. Ecol. 24 (14), 3546–3558. http://dx.doi.org/10.1111/mec.13204, URL: http://doi.wiley.com/10.1111/mec.13204.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38 (6), 1358. http://dx.doi.org/10.2307/2408641, URL: https://www.jstor.org/stable/2408641?origin=crossref.

Wencélius, J., Thomas, M., Barbillon, P., Garine, E., 2016. Interhousehold variability and its effects on seed circulation networks: a case study from northern Cameroon. Ecol. Soc. 21 (1), http://dx.doi.org/10.5751/ES-08208-210144, URL: http://www.ecologyandsociety.org/vol21/iss1/art44/.

Zanke, C.D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, K., Ganal, M.W., Röder, M.S., 2014. Whole genome association mapping of plant height in winter wheat (Triticum aestivum L.). In: Hernandez, P. (Ed.), PLoS One 9 (11), e113287. http://dx.doi.org/10.1371/journal.pone.0113287, URL: https://dx.plos.org/10.1371/journal.pone.0113287.