MOLECULAR ECOLOGY

Molecular Ecology (2014) 23, 2383–2401 doi: 10.1111/mec.12741

INVITED REVIEWS AND SYNTHESES

# An overview of the utility of population simulation software in molecular ecology

SEAN HOBAN

*National Institute for Mathematical and Biological Synthesis, University of Tennessee, 1122 Volunteer Blvd., Suite 110A, Knoxville, TN 37996-3410, USA*

## Abstract

**Stochastic simulation software that simultaneously model genetic, population and environmental processes can inform many topics in molecular ecology. These include forecasting species and community response to environmental change, inferring dispersal ecology, revealing cryptic mating, quantifying past population dynamics, assessing in situ management options and monitoring neutral and adaptive biodiversity change. Advances in population demographic–genetic simulation software, especially with respect to individual life history, landscapes and genetic processes, are transforming and expanding the ways that molecular data can be used. The aim of this review is to explain the roles that such software can play in molecular ecology studies (whether as a principal component or a supporting function) so that researchers can decide whether, when and precisely how simulations can be incorporated into their work. First, I use seven case studies to demonstrate how simulations are employed, their specific advantage/necessity and what alternative or complementary (nonsimulation) approaches are available. I also explain how simulations can be integrated with existing spatial, environmental, historical and genetic data sets. I next describe simulation features that may be of interest to molecular ecologists, such as spatial and behavioural considerations and species' interactions, to provide guidance on how particular simulation capabilities can serve particular needs. Lastly, I discuss the prospect of simulation software in emerging challenges (climate change, biodiversity monitoring, population exploitation) and opportunities (genomics, ancient DNA), in order to emphasize that the scope of simulation-based work is expanding. I also suggest practical considerations, priorities and elements of best practice. This should accelerate the uptake of simulation approaches and firmly embed them as a versatile tool in the molecular ecologist's toolbox.**

*Keywords*: Approximate Bayesian Computation, ex situ management, in situ, landscape, molecular markers, natural history, population dynamics, prediction

*Received 14 November 2013; revision received 22 March 2014; accepted 26 March 2014*

## Introduction

It is an exciting era for computational approaches in ecology and evolutionary biology due to increasing processor power, access to cluster computing facilities, large and increasingly open-access empirical data sets and software whose flexibility and realism makes them usable by a broad community. This is certainly true at the frontiers of molecular ecology, and especially for computer simulations, where software for simulating genetic, environmental and demographic processes are transforming the way molecular data are used to infer and understand ecological processes.

Molecular ecology studies take place at the intersection of population demography, individual behaviour, geography, landscape, history and molecular data, as well as genetic processes such as inbreeding and adaptation. Population demographic and genetic simulators developed over the past several decades are a unique tool for simultaneously modelling these forces across temporal and spatial scales (Caughley 1994; Epperson

Correspondence: Sean Hoban, Fax: 001- (865) 974-9300; E-mail: shoban@alumni.nd.edu

*et al.* 2010; Balkenhol & Landguth 2011), as they allow flexible parameterization of relevant processes (e.g. population sizes, migration, recombination, selection). Parameter-rich and customizable population demographic–genetic simulation software, featuring long-awaited realistic modelling of these relevant processes, are now facilitating a variety of simulation-based investigations at several stages of a study (Supporting Information, Table S1, Fig. 1). They are helping to reveal ecological patterns and process, such as estimating the timing, degree and cause (e.g. exploitation, climate change) of population declines and predecline population sizes (Alter *et al.* 2012), the timing of geneflow cessation (Marino *et al.* 2013) and factors underlying the extent of observed admixture (Perrier *et al.* 2012). Simulations also contribute to elucidating species-specific natural history such as dispersal and mating patterns (Brekke *et al.* 2011; Puebla *et al.* 2012), or landscape barriers. Furthermore, the potential future outcomes of management actions (e.g. translocations), environmental change or captive breeding programmes can be evaluated probabilistically with simulations (Bruford *et al.* 2010). Simulation-based studies have also recently advanced theoretical understanding of range expansion (Travis *et al.* 2007) and retention of adaptive diversity after bottlenecks or fragmentation (Ejsmond & Radwan 2011). Lastly, simulators help evaluate population genetic tools and methods (e.g. estimators of effective population size) by quantifying their performance in real-world conditions (Antao *et al.* 2011; Paz-Vinas *et al.* 2013), and help plan optimal sampling strategies (Gapare *et al.* 2008; Whiteley *et al.* 2012; Hoban *et al.* 2013b). Moreover, simulations can help fully utilize large-scale genetic, geographical, pedigree, historical and ecological data sets, including ancient DNA (Campos *et al.* 2010), and help provide front-line advice and information increasingly sought by conservationists and natural resource policy makers (Cook *et al.* 2013).

Nonetheless, simulations are not yet been widely deployed by molecular ecologists (Andrew *et al.* 2013) probably due to limited accessibility, flexibility and realism of some simulation software and technical challenges in constructing custom-built simulations. Recent user-friendly software are overcoming such limitations, although a guide to their general use and utility is lacking. This review is directed specifically towards molecular ecologists (including ecologists using molecular tools, and population and conservation geneticists), and aims to explain why, where and how simulations can be applied to common study topics. To achieve this aim, I will use detailed case studies, description of valuable simulator capabilities and an outlook on simulation use in emerging areas of molecular ecology. This review is distinct in scope and content from Hoban *et al.* (2012a), which provided: a basic foundation on the topic of genetic simulations; an overview of models of migration, mutation and selection; examples of inference and prediction from ecology, evolutionary biology, epidemiology and anthropology; and comparison of features of 40 available software. This review will focus on features and applications for molecular ecologists and will not discuss simulations for teaching purposes, epidemiology, evolution of DNA architecture or livestock breeding. Readers interested in applications outside molecular ecology or in the technical details of particular software may turn to recent reviews (Carvajal-Rodriguez 2010; Arenas 2012;

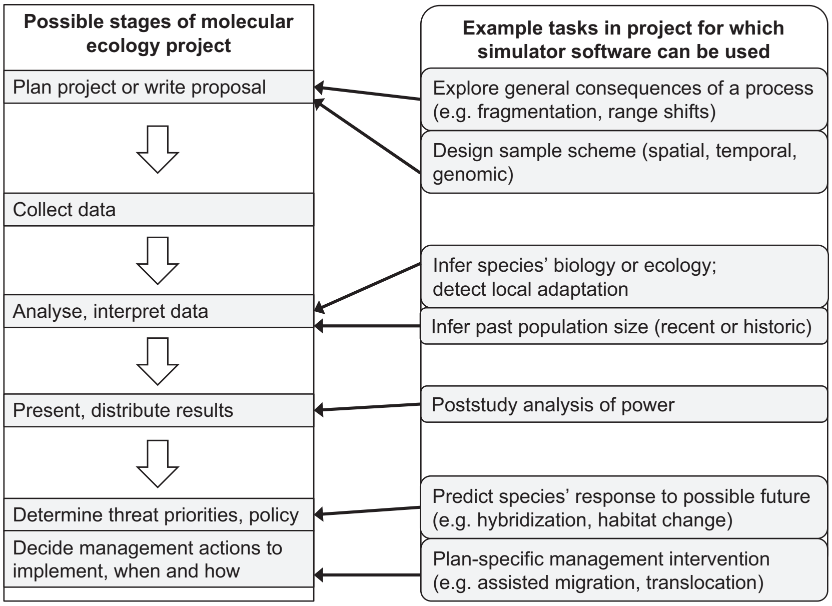

Possible stages of molecular ecology project
Plan project or write proposal
Collect data
Analyse, interpret data
Present, distribute results
Determine threat priorities, policy
Decide management actions to implement, when and how
Example tasks in project for which simulator software can be used
Explore general consequences of a process (e.g. fragmentation, range shifts)
Design sample scheme (spatial, temporal, genomic)
Infer species' biology or ecology; detect local adaptation
Infer past population size (recent or historic)
Poststudy analysis of power
Predict species' response to possible future (e.g. hybridization, habitat change)
Plan-specific management intervention (e.g. assisted migration, translocation)

**Fig. 1** Hypothetical timeline of a molecular ecology project including potential points at which simulations could be employed.

Hoban *et al.* 2012a), or the Genetic Simulation Resources website, http://popmodels.cancercontrol.cancer.gov/gsr/ (Peng *et al.* 2013).

## Case studies

Hoban *et al.* (2012a) suggested three types of simulation-based studies: inferring parameters, predicting potential futures and evaluating methods and models. These broad types may be expanded into seven categories of simulation uses in molecular ecology: (i) infer species' history, (ii) infer species' biology, (iii) predict probable species' response to environmental change (often for policy purposes), (iv) predict outcomes of possible management interventions (for choosing appropriate actions), (v) evaluate method performance, (vi) evaluate/optimize sampling strategy and (vii) explore models of complex processes to discover new phenomena (e.g. emergent properties) or develop theory. This section presents one case study for each category, to demonstrate the use and usefulness of simulations. Many further examples of each type are presented in Table S1 (Supporting Information). I explain what process was of interest, how simulations were used, their specific advantage/necessity and what alternative or complementary (nonsimulation) approaches could be used. In this way, I explain why simulations can sometimes achieve results that cannot be obtained otherwise. I also explain how simulations are sometimes integrated with existing landscape, environmental and genetic data sets (as starting conditions or comparison to simulation outputs), emphasizing that the simulation is only one element involved (summarized in Table 1).

### *Infer species' history*

Population sizes, migration rates and geographical ranges fluctuate throughout a species' history due to ancient (e.g. glacial cycles) and recent (e.g. harvest, habitat fragmentation) events. A challenge for molecular ecologists is to determine how these changes shaped the genetic diversity distributions observed today. Population geneticists long relied on ad hoc interpretation (and overinterpretation) of how summaries of the data (e.g. $F_{ST}$, clustering results) may reflect such histories. Fortunately, techniques for reconstructing demographic history or estimating the influence of environmental variables have arisen in recent years (Marjoram & Tavaré 2006), helping translate abstract genetic patterns into quantitative inference (including measures of uncertainty). Nonetheless, many of these approaches assume particular models (temporally constant migration rates, two populations, etc.) that are rarely met in real populations, sometimes resulting in poor inference. Simulation-based inference (Fig. 2) can utilize models that are arbitrarily complex, allowing inference in dynamic multiple-population situations. To do so, first one defines the model of interest (e.g. five populations, each with different sizes, which may change at some point(s) in the past). Next, one simulates many data sets under this model, with each simulation using different values for key parameters such as dispersal distances and growth rates. Then, the observed data are compared to all the simulated data sets. Parameter inference is made based on parameters of the simulations whose genetic (e.g. heterozygosity, $F_{ST}$) and/or spatial data (dates of colonization) best match the real observations, that is, those parameters that could have plausibly produced the real data. The investigator may use ad hoc methods to compare simulated and observed data in order to bound parameter values (e.g. Mardulyn & Milinkovitch 2005, Mardulyn *et al.* 2009) or may use the formal technique of approximate Bayesian computation (ABC) to compare models, reconstruct the posterior distribution of parameters and determine the reliability of parameter estimation (Bertorelle *et al.* 2010; Csilléry *et al.* 2010). For example, the simulator SPLATCHE2 (Ray *et al.* 2010) was used to construct spatially and temporally explicit, dynamic simulations of two expanding cane toad (*Bufo marinus*) invasions in Australia (Estoup *et al.* 2010). Known or suspected historical dates of founding populations were used as prior information in the simulations. They next compared simulated and observed data to infer number of founders for each establishing population, current population size and number of migrants. Interestingly, the northern invasion, corresponding to more suitable habitat, had more migrants (>100 compared to <5) and larger population sizes than the eastern invasion. These parameters, which deepen our understanding of biological invasions and enhance prediction capabilities, could not have been well estimated using nonsimulation-based approaches because the invasion model is complex and dynamic.

### *Infer species' life history or reproductive biology*

Molecular ecology techniques can help reveal important aspects of natural history, especially in organisms that are difficult to observe directly. An important task is to determine the extent to which organisms disperse from their natal (birth) sites, as dispersal affects rate of inbreeding and metapopulation viability. This is particularly challenging in marine organisms whose larvae are very small and numerous. One molecular technique is to estimate mean axial parent–offspring dispersal distance ($\sigma^2$) from the slope of an isolation-by-distance (IBD) curve (Rousset 1997). However, this technique does not reveal other important aspects of the dispersal kernel, such degree of kurtosis. A recent investigation

**Table 1** Categories of simulation uses in molecular ecology, including examples of tasks performed, processes simulated, complementary data sets and methods, and citations

| Category of simulation study | Specific tasks as well as ecological, anthropogenic or evolutionary processes simulated | Data sets used in tandem with simulations | Examples of alternative or complimentary approaches | Example simulation-based studies |
|---|---|---|---|---|
| Infer history | Estimate historical or current parameter values for processes including population size change, splitting times or colonization. Establish support among alternative histories via model choice | Empirical genetic data required for comparing to simulated data. Historical data useful for priors or for comparison. Landscape data useful in spatially explicit simulations | Bayesian skyline plot or likelihood-based parameter estimation for parameters such as population size, splitting times and migration | Distinguishing between historical and recent secondary contact (Marino *et al.* 2013); inferring invading population size (Ficetola *et al.* 2008; Estoup *et al.* 2010) |
| Infer biology | Estimate parameter values for contemporary processes including variance in reproductive success or migration rates | Empirical genetic data required for comparing to simulated data | Parentage analysis, IBD regression, assignment tests | Inferring dispersal function (Puebla *et al.* 2012) |
| Predict outcome to future environmental changes | Forecast genetic and demographic response to processes including fragmentation, habitat loss, increasing admixture, natural selection or range shifts | Empirical genetic data, present and future landscape (as resistance surface) and historical data optional as starting point for simulations. | Stochastic resampling* for population size changes. Base predictions on observations in related species (e.g. documented past extinctions) | Forecast response to predicted population loss and shifts due to warming (Wasserman *et al.* 2012) |
| Predict response to management | Forecast response to (i) different numbers and timing of translocated individuals, (ii) harvesting of individuals or (iii) ‘no action’ option. Estimate minimum viable genetic population size to avoid genetic erosion | As above. Additionally, full or partial pedigrees in small populations optional | As above. Additionally, meta-analyses of a management action (e.g. translocations) | Translocations and corridors (Bruford 2010); timber harvesting and inbreeding (Degen *et al.* 2006) |
| Evaluate method | Estimate error rates when assumptions of an analytical model are violated. Evaluate performance in extreme conditions, complex situations or for detecting subtle effects/small effect sizes. Compare power of several methods. | Empirical genetic data optional for starting conditions, although usually generic/equilibrium conditions are used | Employ method on population with well-known history, or on laboratory populations subjected to particular histories (fragmentation, bottleneck) | Evaluating methods for detecting bottlenecks (Chikhi *et al.* 2010), local adaptation (De Mita *et al.* 2013); barriers (Blair *et al.* 2012) |
| Evaluate potential sample strategies | Determine minimum number of markers or samples to achieve desired level of power. Determine best marker type. | As above | Utilize sampling strategy that was successful in related species/situation | Optimize sample strategy to estimate Ne (Antao *et al.* 2011) |

**Table 1** *Continued*

| Category of simulation study | Specific tasks as well as ecological, anthropogenic or evolutionary processes simulated | Data sets used in tandem with simulations | Examples of alternative or complimentary approaches | Example simulation-based studies |
|---|---|---|---|---|
| | Compare 'more markers' vs. 'more samples' | | | |
| Explore process/ model | Simulate complex process of interest (e.g. range expansion with selection). Explore consequences of species' life history traits | As above | Mathematical or conceptual models | Understand neutral and adaptive changes during range expansion (Peischl *et al.* 2013); understand how breeding system attenuates bottleneck effects (Hundertmark & Van Daele 2010) |

*Although resampling techniques are sometimes considered a type of simulation.

of dispersal in five coral reef fish species (Puebla *et al.* 2012) employed the IBD technique, but complemented this with simulations to infer other aspects of the dispersal kernel (Fig. 2). Using the individual-based simulator IBDSIM (Leblois *et al.* 2009), they created populations whose density and geometric arrangement matched the real study population. They simulated a variety of dispersal kernels and recorded the IBD slope generated by each simulation. The idea was to search for a 'match' between the IBD pattern observed in the real population genetic data and the IBD patterns generated by the various simulated dispersal kernels; 'matches' identified kernels that could have plausibly generated the observed data. As noted above, comparison of simulated and observed data may be relatively informal, for example observed values are within 95% of the simulated data, as in this case study, or formally through the ABC procedure. Interestingly, species' dispersal distance (determined by both methods) correlates with duration of their pelagic larval stage, a finding that informs decisions about the size of marine management areas. This simulation-based technique for identifying plausible dispersal kernels can complement the IBD technique or new Bayesian methods (Moran & Clark 2011). The simulation approach has two advantages: the investigator (i) can test a wide variety of dispersal kernels (including complex kernels) and (ii) can create simulations whose density, geometric arrangement and sampling scheme are similar to the real population, which obviates assumptions implicit to the IBD approach.

### *Predict probable species' response to environmental change*

Prediction of species' response to environmental change is necessary for informing population prioritization and protection, but is a difficult challenge because populations experience multiple ongoing processes: climate warming, fragmentation, reduction in habitat area and population size decline. Ideally, predictions need to be spatially explicit and for a range of possible climate scenarios. Fortunately, simulations of populations subject to relevant environmental influences can be used to forecast likely outcomes (Fig. 3). The idea is to observe the in silico population dynamics (e.g. extinction, range contraction) and use these as a basis for informed forecasting of plausible real-world dynamics. A study of American marten (*Martes americana*) used simulations on realistic landscapes representing the northern Rocky Mountains to determine how number of alleles and heterozygosity could respond to reductions in suitable habitat caused by climate change (specifically, an increase in optimal altitude due to warming). Wasserman *et al.*

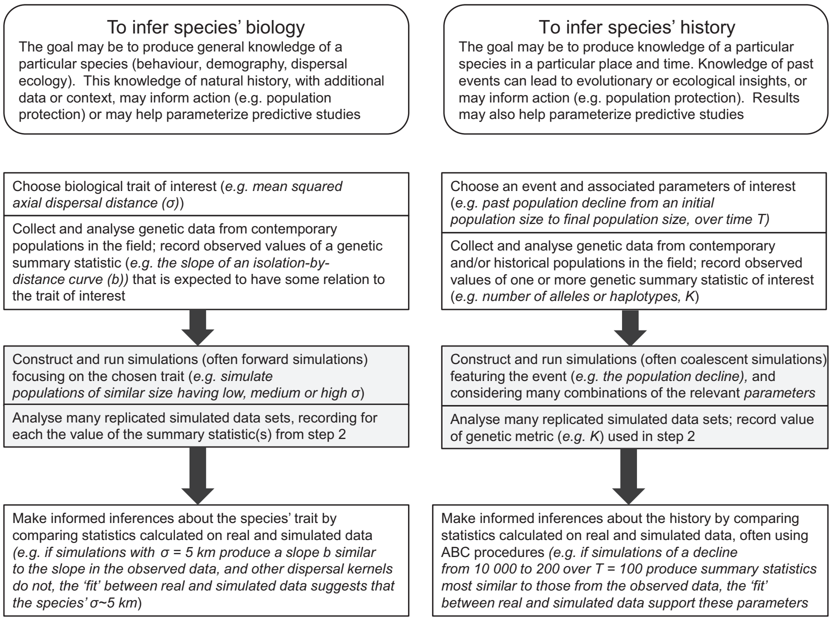

To infer species' biology
The goal may be to produce general knowledge of a particular species (behaviour, demography, dispersal ecology). This knowledge of natural history, with additional data or context, may inform action (e.g. population protection) or may help parameterize predictive studies
Choose biological trait of interest (e.g. mean squared axial dispersal distance (σ))
Collect and analyse genetic data from contemporary populations in the field; record observed values of a genetic summary statistic (e.g. the slope of an isolation-by-distance curve (b)) that is expected to have some relation to the trait of interest
Construct and run simulations (often forward simulations) focusing on the chosen trait (e.g. simulate populations of similar size having low, medium or high σ)
Analyse many replicated simulated data sets, recording for each the value of the summary statistic(s) from step 2
Make informed inferences about the species' trait by comparing statistics calculated on real and simulated data (e.g. if simulations with σ = 5 km produce a slope b similar to the slope in the observed data, and other dispersal kernels do not, the 'fit' between real and simulated data suggests that the species' σ~5 km)
To infer species' history
The goal may be to produce knowledge of a particular species in a particular place and time. Knowledge of past events can lead to evolutionary or ecological insights, or may inform action (e.g. population protection). Results may also help parameterize predictive studies
Choose an event and associated parameters of interest (e.g. past population decline from an initial population size to final population size, over time T)
Collect and analyse genetic data from contemporary and/or historical populations in the field; record observed values of one or more genetic summary statistic of interest (e.g. number of alleles or haplotypes, K)
Construct and run simulations (often coalescent simulations) featuring the event (e.g. the population decline), and considering many combinations of the relevant parameters
Analyse many replicated simulated data sets; record value of genetic metric (e.g. K) used in step 2
Make informed inferences about the history by comparing statistics calculated on real and simulated data, often using ABC procedures (e.g. if simulations of a decline from 10 000 to 200 over T = 100 produce summary statistics most similar to those from the observed data, the 'fit' between real and simulated data support these parameters

**Fig. 2** Flowcharts of possible steps involved in hypothetical simulation-based studies for two types of inference (to be used as a guide rather than as prescriptive).

(2012) used the individual-based software CDPOP (Landguth & Cushman 2010) to identify regions in which climate would cause high losses (peripheral populations) and regions that would be little affected. Altitude and local habitat characteristics were incorporated into the model of individual dispersal and persistence. They also observed a nonlinear response of heterozygosity, suggesting thresholds of fragmentation past which large genetic changes will occur. This result highlights another particular advantage of simulations in this context: fine adjustment of parameters can help identify thresholds (critical levels that are 'tipping points'). The results of this study were then extrapolated to predict genetic dynamics over a much larger area of the Rockies. Because individual-based simulations over this large area would be computationally intractable, they used regressions of genetic responses (heterozygosity and number of alleles) on predictor variables (largest patch size, density of patches, aggregation). This combined methodology (simulation at medium-scale plus extrapolation) helped identify regionally important corridors and barriers. Another approach for predicting regions that may experience gene loss is species distribution models (SDM), which may help identify populations in habitat that will become unsuitable. However, simulations allow more realistic and spatially and temporally explicit prediction of demographic and genetic response. Individual-based simulations are ideal for this category of investigation, but are usually limited to small scales. Thus, simulations must be combined with clever approaches like the regression technique for making regional- or continental-scale predictions. This category of spatial simulation requires a cost-distance landscape, usually produced with genetic data and least-cost path modelling, which may not be available in some species; SDM predictions require different data (large-scale presence–absence records).

### *Predict outcomes of possible management interventions to choose appropriate action*

Choosing suitable management options is another applied challenge for molecular ecology. Appropriate management actions should result in a desirable genetic and demographic outcome, such as retention of a given proportion of heterozygosity, maintenance of high effective population size or reduction in unwanted admixture. Meta-analyses of management attempts can help provide general guidance (Godefroid *et al.* 2011), but it is desirable to optimize management to particular species and locations. One can use simulations of populations subject to various potential actions (translocation, re-introduction, culling) to quantify probable outcomes, revealing which actions can result in desirable outcomes (Fig. 3). (The potential actions are defined by the resources of managers, and situation-specific social and political constraints.) An advantage of simulations is that the many repeated simulation iterations from the same initial conditions (termed Monte Carlo simulation) help quantify possible variation and thus uncertainty in outcome, a key aspect of management. Brekke *et al.* (2011) used BOTTLESIM (Kuo & Janzen 2003) to create individual-based simulations tailored to observed sex ratios and known source population size

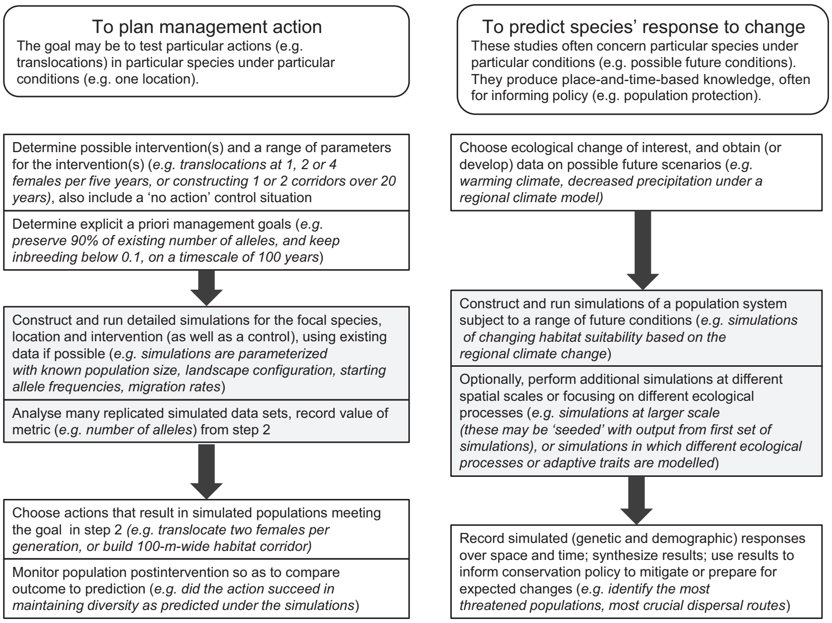

To plan management action
The goal may be to test particular actions (e.g. translocations) in particular species under particular conditions (e.g. one location).
Determine possible intervention(s) and a range of parameters for the intervention(s) (e.g. translocations at 1, 2 or 4 females per five years, or constructing 1 or 2 corridors over 20 years), also include a 'no action' control situation
Determine explicit a priori management goals (e.g. preserve 90% of existing number of alleles, and keep inbreeding below 0.1, on a timescale of 100 years)
Construct and run detailed simulations for the focal species, location and intervention (as well as a control), using existing data if possible (e.g. simulations are parameterized with known population size, landscape configuration, starting allele frequencies, migration rates)
Analyse many replicated simulated data sets, record value of metric (e.g. number of alleles) from step 2
Choose actions that result in simulated populations meeting the goal in step 2 (e.g. translocate two females per generation, or build 100-m-wide habitat corridor)
Monitor population postintervention so as to compare outcome to prediction (e.g. did the action succeed in maintaining diversity as predicted under the simulations)
To predict species' response to change
These studies often concern particular species under particular conditions (e.g. possible future conditions). They produce place-and-time-based knowledge, often for informing policy (e.g. population protection).
Choose ecological change of interest, and obtain (or develop) data on possible future scenarios (e.g. warming climate, decreased precipitation under a regional climate model)
Construct and run simulations of a population system subject to a range of future conditions (e.g. simulations of changing habitat suitability based on the regional climate change)
Optionally, perform additional simulations at different spatial scales or focusing on different ecological processes (e.g. simulations at larger scale (these may be 'seeded' with output from first set of simulations), or simulations in which different ecological processes or adaptive traits are modelled)
Record simulated (genetic and demographic) responses over space and time; synthesize results; use results to inform conservation policy to mitigate or prepare for expected changes (e.g. identify the most threatened populations, most crucial dispersal routes)

**Fig. 3** Flowcharts of possible steps involved in hypothetical simulation-based studies for two types of prediction (to be used as a guide rather than as prescriptive).

to determine appropriate introduction sizes for hihi (*Notiomystis cincta*), an endangered New Zealand bird, on predator-free islands. They also tested how different assumed growth rates and different mating strategies, including polygony (single male dominance) and partial dominance by a single pair, would affect genetic diversity. [This process – quantifying the effects of including different model components and testing different values for key parameters – is termed sensitivity analysis. This crucial component of simulation-based studies helps an investigator to assign variation in model output to particular model inputs, and to identify the most influential parameters (Peck 2004; Naujokaitis-Lewis *et al.* 2009).] Simulations showed that the ideal number to introduce depended highly on growth rate and mating strategy. This result suggests that managers must seek to maximize postrelease growth (which may require choosing sites with large carrying capacities and/or food supplementation) and must consider mating opportunities. Simple stochastic resampling techniques are an alternative approach to predicting genetic diversity loss during introductions, but this technique cannot represent demographic stochasticity or realistic mating systems. Nonetheless, Brekke *et al.* used resampling in another part of their investigation to estimate that approximately half of genetic diversity loss occurs immediately after introduction, and half over time through drift, again emphasizing rapid growth rates. Another alternative is custom-built simulation models, which are highly detailed but difficult to create. Available, flexible simulation software are a balance between the simplicity of resampling and the complexity and specificity of custom simulations.

### *Evaluate method performance*

Molecular ecology is experiencing a time of rapid development and uptake of new statistical methods. However, new methods and tools are typically evaluated by their authors only in simple, general conditions. Unfortunately, some methods may identify false signals (type I errors) or otherwise show poor performance when applied under real-world conditions, especially when a method relies on simple theoretical models. Rigorous method evaluation is thus an imperative and vital contribution to the field, to advise when methods should and should not be used (Fig. 4). For example, several bottleneck detection methods (commonly used in molecular ecology) assume an isolated (no migration) population of historically constant size reduced instantaneously to a new constant (smaller) size. Life history is assumed to be simple, and a mutation model is defined. However, real-world situations rarely conform to this assumed model. Connectivity to other populations is one violation of the assumptions. Two investigations recently evaluated bottleneck detection methods by constructing simulations [with the software MS, (Hudson 2002)] in which bottlenecks did or did not occur and where populations were connected by varying degrees of migration. For each simulated data set, they applied the bottleneck test(s) and reached a conclusion based on the test results (e.g. a bottleneck did occur). Because they used simulated data, they knew whether the simulated population actually did experience a bottleneck, so they could quantify how often the method made erroneous conclusions. Chikhi *et al.* (2010) showed that migration

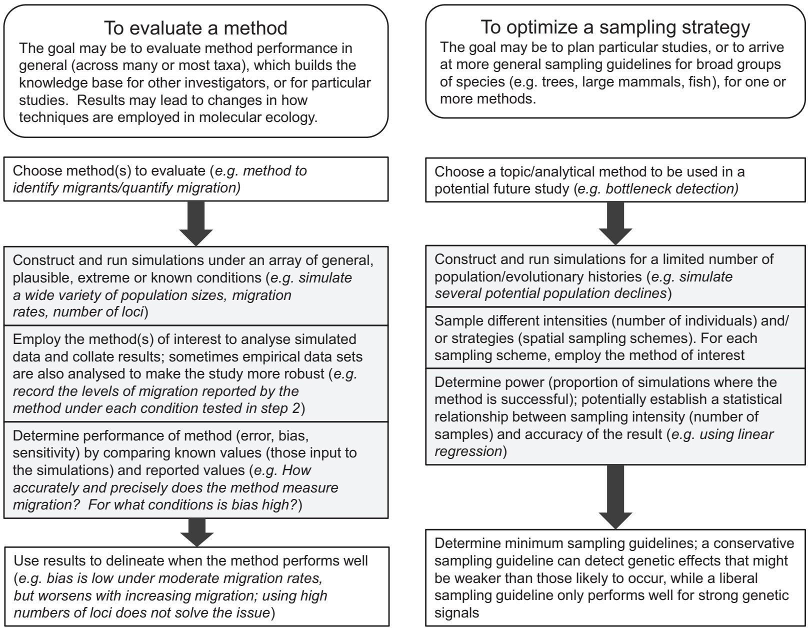
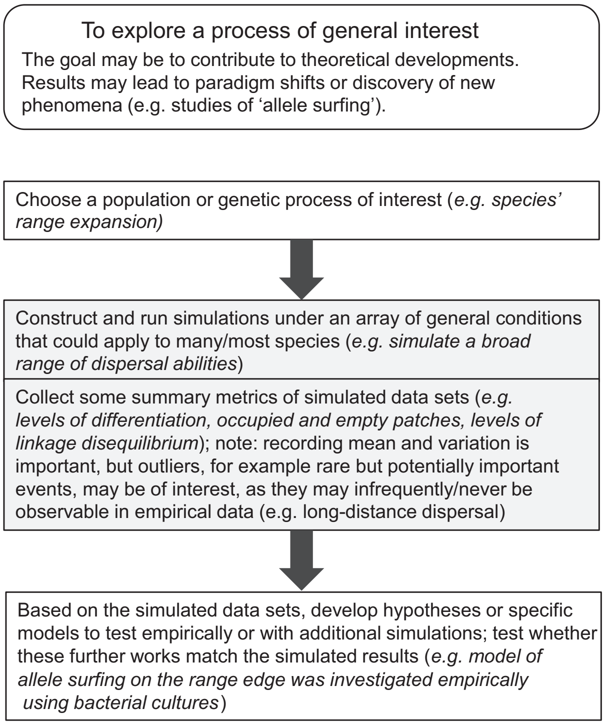

To evaluate a method
The goal may be to evaluate method performance in general (across many or most taxa), which builds the knowledge base for other investigators, or for particular studies. Results may lead to changes in how techniques are employed in molecular ecology.
Choose method(s) to evaluate (e.g. method to identify migrants/quantify migration)
Construct and run simulations under an array of general, plausible, extreme or known conditions (e.g. simulate a wide variety of population sizes, migration rates, number of loci)
Employ the method(s) of interest to analyse simulated data and collate results; sometimes empirical data sets are also analysed to make the study more robust (e.g. record the levels of migration reported by the method under each condition tested in step 2)
Determine performance of method (error, bias, sensitivity) by comparing known values (those input to the simulations) and reported values (e.g. How accurately and precisely does the method measure migration? For what conditions is bias high?)
Use results to delineate when the method performs well (e.g. bias is low under moderate migration rates, but worsens with increasing migration; using high numbers of loci does not solve the issue)
To optimize a sampling strategy
The goal may be to plan particular studies, or to arrive at more general sampling guidelines for broad groups of species (e.g. trees, large mammals, fish), for one or more methods.
Choose a topic/analytical method to be used in a potential future study (e.g. bottleneck detection)
Construct and run simulations for a limited number of population/evolutionary histories (e.g. simulate several potential population declines)
Sample different intensities (number of individuals) and/or strategies (spatial sampling schemes). For each sampling scheme, employ the method of interest
Determine power (proportion of simulations where the method is successful); potentially establish a statistical relationship between sampling intensity (number of samples) and accuracy of the result (e.g. using linear regression)
Determine minimum sampling guidelines; a conservative sampling guideline can detect genetic effects that might be weaker than those likely to occur, while a liberal sampling guideline only performs well for strong genetic signals

**Fig. 4** Flowcharts of possible steps involved in hypothetical simulation-based studies for evaluation and optimization (to be used as a guide rather than as prescriptive).

between genetically differentiated populations will cause false bottleneck signals (because migrants introduce rare allelic variants), while Paz-Vinas *et al.* (2013) showed that asymmetrical migration (higher in one direction, e.g. river populations) may trigger false expansion signals. These studies emphasize caution when interpreting bottleneck tests and suggest that simulations should accompany studies using these methods. Another way to evaluate a method is to apply it to empirical populations with known history (e.g. laboratory populations subjected to bottlenecks), but such an approach is obviously limited in number of replications and parameter values that can be tested (England *et al.* 2003). Furthermore, with simulations, the investigator specifies all starting conditions, so the processes producing the data are known with certainty, allowing explicit attribution of a result to its cause (which is rare for empirical systems). However, empirical-based evaluations sometimes have the advantage of realism; the simulation approach assumes that all important characteristics of the real world are incorporated into the simulation model, which may not always be true. Indeed, early evaluations of bottleneck methods did not consider connectivity. There is ongoing need for simulation-based studies to quantify how organism attributes (overlapping generations, mating systems), within-population structure (kin groups, spatial structure) and recent demographic history may adversely affect performance of commonly employed methods, such as clustering, assignment, outlier detection, *Ne* estimation and isolation-by-resistance methods (Whiteley *et al.* 2012; Hoban *et al.* 2013d; De Mita *et al.* 2013).

### *Evaluate/optimize sampling strategy*

It is important that researchers utilize sampling/monitoring protocols that are well designed and robust, in order to ensure appropriate collection, analysis and interpretation of increasingly large and complex genetic data sets. One key task is to collect sufficient samples (number of observations) to ensure statistically significant and/or precise results (detection of an effect of interest), while also being efficient, as funding and time is limited. It is also increasingly important to decide what marker type best suits an investigation (SNPs, microsatellites, NGS-obtained sequences). Unfortunately, many genetic studies are undertaken without knowledge of how many populations, genetic markers or individuals are needed. The appropriate sampling strategy depends on the task (i.e. different sampling is needed for landscape genetics, local adaptation, identifying hybrids, detecting bottlenecks, etc.) and species. For this reason, perhaps the most common nonsimulation approach to planning a study (i.e. utilize sampling strategy that was successful in a different study) is not optimal (Hoban *et al.* 2013b). Sampling strategy can be optimized by quantifying the statistical power (approximately the probability of success) of all potential sampling strategies and by choosing one that results in acceptable power (e.g. >0.90). Simulations are used to estimate power (Fig. 4) as follows (see Hoban *et al.* 2013b). First, one constructs models of population/evolutionary histories (say, a small, medium and severe bottleneck) representing the real world and chooses a range of plausible sampling strategies. For each model, one: (i) runs a simulation, (ii) collects individual simulated genotypes, (iii) using this genetic data, performs the analyses that would be performed in the planned study, (iv) reaches a conclusion from this test (e.g. a bottleneck did occur) and (v) quantifies how well that sampling performed (is the conclusion correct, how 'far' from the true value – the one used to simulate the data – is the estimate). Based on many replicates, one obtains an estimate of the relative power of each sampling strategy to detect particular genetic effects in populations approximately similar to those simulated. Landguth *et al.* (2012) used this approach to optimize sampling for causal modelling to distinguish isolation-by-resistance from other hypotheses, in black bear (*Ursus americanus*). Using the steps outlined above, they demonstrated that 300–500 individuals and 15 loci provide sufficient power in their situation; furthermore, a moderate increase to 25 markers could reduce the necessary individuals to 100. This utility of applying more markers is an important consideration because genotyping costs are constantly decreasing and because many populations are small (<100 individuals). It should be noted that this strategy is suitable for their particular study; individual investigators (or funding agencies) can and should use simulators to custom-design sampling strategies to their needs before beginning studies. Notable, in some cases, genetic effects are undetectable with any reasonable sampling and such studies should not proceed (Hoban *et al.* 2013a,b). General guidelines (e.g. 5–20 polymorphic loci and 20–30 individuals for one bottleneck test) can be developed by simulating general or intermediate conditions (Cornuet & Luikart 1996).



To explore a process of general interest

The goal may be to contribute to theoretical developments. Results may lead to paradigm shifts or discovery of new phenomena (e.g. studies of 'allele surfing').

Choose a population or genetic process of interest (*e.g. species' range expansion)*

Construct and run simulations under an array of general conditions that could apply to many/most species (*e.g. simulate a broad range of dispersal abilities*)

Collect some summary metrics of simulated data sets (*e.g. levels of differentiation, occupied and empty patches, levels of linkage disequilibrium*); note: recording mean and variation is important, but outliers, for example rare but potentially important events, may be of interest, as they may infrequently/never be observable in empirical data (e.g. long-distance dispersal)

Based on the simulated data sets, develop hypotheses or specific models to test empirically or with additional simulations; test whether these further works match the simulated results (*e.g. model of allele surfing on the range edge was investigated empirically using bacterial cultures*)

**Fig. 5** Flowcharts of possible steps involved in hypothetical simulation-based studies for theory development or process exploration (to be used as a guide rather than as prescriptive).

### *Explore complex processes to garner theoretical insight*

Various ecological, spatial, demographic, genetic and anthropogenic processes (e.g. population bottlenecks, range shifts, selection) may have complex consequences for the amount, type and distribution of neutral and adaptive genetic diversity (between and within populations). Additionally, such processes when combined may have nonlinear, interacting and/or compensatory effects. While empirical studies (experimental populations or well-studied natural populations) in aggregate help contribute towards theoretical understanding of complex processes (e.g. Petit *et al.* 2002), replication is often low, and rare events may not be detected. Simulations provide an alternative avenue for theoretical advances (Fig. 5) and have recently advanced our general understanding of the intersection of ecology and evolution. Indeed, simulations have a long history in theoretical studies, such as to verify analytical results or explore complex, analytically intractable situations, producing major advances in population genetics long before modern molecular techniques (Kimura *et al.* 1975; Maruyama & Kimura 1980). The idea is that an investigator can simulate a process of interest over different temporal and spatial scales and observe the genetic diversity and distribution outcomes; if simulations are realistic, the observed dynamics may represent those occurring in some real populations or species. For example, some molecular ecologists seek to understand how potentially beneficial climate-adapted alleles spread among populations and more broadly to understand species' adaptive potential. Schiffers *et al.* (2013) used simulations of a hypothetical species with quantitative trait loci adapted to a climatic gradient and separate loci adapted to local environs (e.g. edaphic conditions). The simulated populations (with the software ALADYN, http://www.katja-schiffers.eu/docs/Aladyn.zip) were then subject to gradual warming; results showed that climate alleles can easily spread throughout a species' range when local adaptation is absent; however, the presence of local adaptation resulted in poor survival of new migrants, preventing them from contributing alleles and slowing the spread of climate adaptations. The advantage of simulators in such work is that processes can be treated in isolation or combined, helping to identify possible interactions of multiple processes, including nonadditive effects and system-level outcomes. Another advantage of exploratory simulations is that they can reveal outliers – infrequent but potentially remarkable events that rarely are observed in real systems (Peck 2004), such as long-distance dispersal. One disadvantage to simulation-based explorations is that findings may somewhat depend on model specifics like genetic architecture of simulated traits. However, several possibilities can be modelled. For example, Schiffers *et al*. tested various models of linkage. Many topics are amenable to simulation-based exploration, including biological invasions, spatial range expansions, fragmentation and retention of adaptive diversity during bottlenecks.

## New simulation software features

Advances in simulation software greatly facilitate studies similar to those in the previous section. These advances include enhanced biological, ecological and genetic models, enabling the user to consider a diversity of processes (Table 1).

### *Landscapes*

Molecular ecologists are often interested in identifying migrants, quantifying the connectivity consequences of natural and human-made barriers, understanding the outcomes of spatially shifting ranges, predicting metapopulation dynamics and detecting local adaptation (Landguth *et al.* 2010; Epperson *et al.* 2010). These tasks are facilitated by simulators that model heterogeneous landscapes that can change over time, and which are occupied by individuals who mate, reproduce and disperse in continuous space (e.g. in CDPOP, KERNELPOP and IBDSIM). A key advance was to provide each individual with unique x-y coordinates. This modelling approach is in contrast to the 'discrete population' paradigm that has until recently dominated molecular ecology, in which individuals occur only in distinct localities connected by migration. It was also important to model movement (adults seeking mates and juveniles dispersing) according to realistic dispersal kernels (e.g. exponential, power), as in the coral reef fish study, rather than as migration rates between populations. For many species, this better represents spatial distribution and allows the development of spatial genetic structure (Strand & Niehaus 2007; Gapare *et al.* 2008; Landguth *et al.* 2012). It must be noted that different data are needed as input for 'discrete population' and 'continuous' models; a migration matrix (stating the proportion of each population exchanging migrants to/from every other population) is used in the former case, while defined dispersal kernels for individual movements are used in the latter case. (In one software, SPLATCHE2, a demographic simulation is used to create the necessary migration matrices.) Another advance is to allow local habitat conditions like forest cover to affect survival and/or dispersal of individuals, and to allow temporal change in environment (e.g. the case study of American marten), as implemented in SPLATCHE2 and CDPOP. GIS data on topography, rivers and vegetation can be loaded into the simulation starting conditions (Landguth & Cushman 2010). This allows testing of how particular ecological or geographical features (forest, prairie, rivers) interact with demographic dynamics (population growth, density) to influence genetic variation and population distribution outcomes.

Example studies with these features include quantifying rates of dispersal across large networks of marine habitat (Alberto *et al.* 2010), or predicting consequences of infrastructure development, such as proposed roads or pipelines. Simulations incorporating realistic landscapes can also help design optimal conservation corridors, predict how populations will respond to habitat changes and determine whether responses will be detectable (Segelbacher *et al.* 2010), also as in the study of American marten (Wasserman *et al.* 2013). Other recent studies have explored the capabilities of species to range shift in response to climate change (naturally or with assistance), and the consequences of such shifts. For example, simulations on a realistic map of Europe (including altitude and water barriers) explored range expansion outcomes for deleterious, beneficial and neutral alleles and found an important effect of mountainous barriers (Travis *et al.* 2007). Simulations also suggested that well-adapted alleles from the centre of the range may have difficulty moving north. Of course, parameterizing realistic landscape genetics simulations comes with challenges, and recent reviews offer suggestions and cautions (Spear *et al.* 2010; Segelbacher *et al.* 2010), such as distinguishing between gene flow and movement, and weighing the benefits of multivariate vs. univariate surfaces.

### *Life history*

A simple population simulation might feature hermaphroditic individuals that mate at random, produce a number of offspring drawn from a Poisson distribution and all die together (semelparous life history). However, this may be a poor model for studying genetic dynamics of real populations (Strand & Niehaus 2007; Lloyd *et al.* 2013). New simulators feature various mating systems (e.g. monogamy, partial or complete dominance by one adult) and offspring distributions (e.g. constant number, skewed reproductive success, age-dependent offspring production), and often allow multiple age classes (progress to the next stage is usually based on probability matrices), which can lead to overlapping generations and possibly intergenerational mating. Age or stage structure may be important to consider as this aspect influences effective population sizes, competition and adaptation under environmental conditions having substantial interannual variation. Sex ratios may be biased, and dispersal, age at maturity and other attributes may differ between sexes. Software in this realm include RMETASIM, CDPOP, NEMO and KERNELPOP.

One area in which these life history features may be crucial is biological invasions, which can cause large ecological and evolutionary change. There is current interest in estimating parameters of invading populations (especially for monitoring and eradication efforts) such as number and source of founding invaders, dispersal, growth rates and timing of introduction (Estoup *et al.* 2010). Realistic simulations are key to providing this information. For example, Ficetola *et al.* (2008) used simulations with the software RMETASIM, http://cran.r-project.org/web/packages/rmetasim/, which allows age-specific mating and dispersal probabilities, for each sex, to infer that invading populations of American bullfrog (*Rana catesbeiana*) in France and Italy stemmed from small introductions (two to six females). Their finding poses a challenge for detecting and eradicating this invader, as new invasions could occur if few individuals escape or are intentionally released. Another useful area for realistic life history simulations is endangered species' re-introductions, as with the hihi case study (Brekke *et al.* 2011).

Life history also influences the evolutionary impact of population decline, such that bottleneck effects are seen in some cases where they are not expected, and vice versa. Recent simulations have been used to explore how mating patterns (e.g. monogamy, lek, extra-pair mating) and overlapping generations interact with demographic decline or expansion to influence diversity patterns. For example, simulations using BOTTLESIM showed that long generation time in the case of DDT-induced decline in eagles (*Haliaeetus albicilla*) or growth rate in the case of introduced elk (*Cervus elaphus roosevelti*) can explain the absence of a genetic bottleneck signal in known demographic bottlenecks (Hailer *et al.* 2006; Hundertmark & Van Daele 2010). The later study used simulations of populations with various growth rates to show that rapid post-bottleneck recovery likely tempered genetic bottleneck signals in spite of an introduction size of only eight animals (current size >1400), important knowledge for captive breeding and re-introductions.

## *Interspecific interactions*

Another realistic improvement is the ability to simulate multiple species simultaneously, including potential interactions such as competition, facilitation or hybridization. For example, multiple-taxa simulations can help determine what rates of interspecific hybridization could have produced currently observed admixture (Perrier *et al.* 2012), or infer the selection strength and genetic architecture of potential barriers to reproductive isolation (e.g. phenological separation). Currat & Excoffier (2011) investigated the long-standing question of interactions between Neanderthals and humans; simulations made with SPLATCHE2 (Ray *et al.* 2010), featuring both density-dependent competition and admixture, allowed the authors to infer a low level of successful interbreeding (<2%). As yet, few simulators feature species' interactions, although NEMO (Guillaume & Rougemont 2006) implements within-host dynamics of Wolbachia, an endosymbiont parasite. Other uses for this feature were highlighted in a recent 'roadmap' for molecular ecology (Andrew *et al.* 2013): in community genomics, such as determining the degree to which genetic variation in one species contributes to ecological processes or genetic diversity of other species (Schweitzer *et al.* 2004), or community phylogeography, for investigating multiple species' simultaneous response to climate warming (e.g. shared postglacial responses, exemplified by Carnaval *et al.* 2009).

## *Adaptive traits and genomics*

Molecular ecologists and conservation biologists are increasingly attentive to contemporary evolutionary change (Stockwell *et al.* 2012). Selection pressures from harvest, exposure to new environments (degraded habitat, captivity) or interactions with exotic species can cause shifts in distributions of traits, for example size. Thus, a valuable feature is neutral or selected phenotypic traits under potentially complex genetic control, allowing the investigation of combined effects of selection, mutation and recombination in multiple-population systems with potentially differing local selection pressures. One approach to modelling traits is more phenotypic [e.g. the software PEDAGOG, Coombs *et al.* 2010)], in which the user defines heritability, type (stabilizing, etc.) and strength of selection, with assumed simple genetic architecture (additive model). The other approach is more genotypic (e.g. the software QUANTINEMO Neuenschwander *et al.* 2008), in which the investigator defines the genetic architecture of traits, including number of loci; whether effects are additive, dominant or epistatic; and environmental effects. It is possible to define chromosomal structure (distance between loci). These features will likely contribute to an ongoing interest in molecular ecology and evolutionary biology: determining how many loci contribute to phenotypic traits, what is the effect size of each and how rapidly can adaptation occur (Yeaman & Guillaume 2009; Yeaman & Whitlock 2011). It may be important to examine multiple traits simultaneously, as a given environmental change may affect multiple characters, with possibly opposing fitness consequences (Shaw & Etterson 2012). The case study of Schiffers *et al.* (2013) is one example of simulating adaptive response to local and global selective pressures, specifically adaptation to local, unchanging conditions (e.g. soil, herbivores) and to changing conditions (climate). Simulations featuring realistic models of adaptation are also necessary for evaluating power of new methods to detect loci under selection, as in a recent study that found that genotype–environment correlation methods had higher power (but also more false positives) than FST outlier methods (De Mita *et al.* 2013).

An increasing number of molecular ecology and population genetic studies utilize thousands of genetic markers, for a variety of purposes but especially for detecting signatures of adaptation. A small number of simulation software [e.g. SFS_code, (Hernandez 2008) and SLIM (Messer 2013)] have specialized in new algorithms and highly optimized code to allow simulation of genome-scale data, including tens of thousands of markers or long (>1 MB) sequences. Simulations of genomic data sets not subject to selection can be used as a null distribution for identifying outlier loci (Gossmann *et al.* 2010), while simulations of different strengths of selection and different degrees of recombination can help infer what type of selection likely produced patterns observed in an empirical data set (Lohmueller *et al.* 2011). A recent study used genomic simulations to quantify the expected frequency of rare variants segregating in an expanding population (i.e. the very recent human population explosion) and to estimate how many rare variants might actually be identified with the sample sizes typically used in genome-wide association studies of human diseases (Keinan & Clark 2012). Note that, as yet, there is typically a trade-off in biological or spatial complexity such that genomic-scale simulations usually utilize simple demographic and life history models.

### *Movement behaviour*

A final improvement regards organism movements. Landscape genetics has contributed to our understanding of connectivity often by modelling movement to minimize resistance [e.g. 'least costs paths (LCP)']. However, LCP assumes that individual organisms have 'omniscience' of the total landscape, and unfailingly choose the optimal habitat for their next step(s), which are typically untrue (Spear *et al.* 2010). Recent simulation software therefore model movement as a series of behavioural decisions (Palmer *et al.* 2011; Rebaudo *et al.* 2013). Organisms may have limited visual perception (e.g. hundreds of metres rather than several kilometres), may choose their next step based on probabilities (e.g. 0.6 probability of turning left and 0.4 probability of turning right, based on slightly better habitat quality to the left) and may show disposition towards straight-line movement (i.e. directional tendency, whereby sharp turns are avoided). Simulations accounting for these factors have shown: that organisms are unlikely to follow LCPs; that mean cost to travel between patches is much higher than LCPs; and that if the number of 'moves' is limited, organisms may not reach a nearby optimal patch, even if they could do so along the LCP (Palmer *et al.* 2011).

## Simulators in emerging frontiers

Molecular ecologists are increasingly challenged to understand and explain the consequences of complex environmental changes, but also possess large data sets unimagined even a few years ago. Population demographic–genetic simulation software can assist in climate change mitigation, biodiversity monitoring and intensive management of exploited populations. New simulation software will also help exploit the latest advances in genomics and historic/ancient DNA.

Simulations can be combined with climate, weather or hydrological models to predict outcomes such as range dynamics and rapid evolutionary changes. In a prime example of using simulations to explore processes at multiple scales, Kramer *et al.* (2010) simulated demographic and genetic response of beech (*Fagus sylvatica*) to predicted temperature and rainfall changes. The authors also investigated response in quantitative traits (bud-burst, drought tolerance) by combining an individual-based genetic model with ecological models of soil absorption. This was performed for leading-edge (selection for earlier bud-burst) and trailing-edge (selection for water-use efficiency) conditions. While neutral diversity was largely unaffected (due to gene flow and large population size), substantial response occurred for quantitative traits, suggesting a potential for rapid adaptation. These authors also simulated management actions such as particular silvicultural practices and artificial gene flow and showed that these interventions could assist adaptation. In another high-impact study, Banks *et al.* (2010) investigated connectivity between recently established, poleward-expanding populations of the invasive, ecosystem-altering sea urchin (*Centrostephanus rodgersii*). They inferred substantial, ongoing migration, implying that newly established and source populations are well connected and suggesting that urchins have rapidly responded to warming ocean temperatures with range expansion into previously unsuitable habitat.

Recently, the Convention on Biological Diversity generated 20 targets for preserving biodiversity by 2020 (http://www.cbd.int/sp/targets/). Simulations will be central for evaluating indicators used for monitoring progress towards these and other conservation targets, for example various measures of genetic, species and ecosystem change (Hoban *et al.* 2013c; Pereira *et al.* 2013). Simulations will help choose appropriate indicators (those that accurately and quickly respond to change), and plan sampling strategies to ensure that temporally collected observations are sufficient (Hoban *et al.* 2013b). Predictive simulations can also determine whether an intervention will likely meet a target, to help weigh policy options. Lastly and importantly, trade-offs and synergies may exist among targets (Perrings *et al.* 2010); simulations could facilitate predicting simultaneous effects of interventions on multiple targets, such as the effect of protected areas (Target 11) or sustainable forestry (Target 7) on genetic diversity (Target 13), extinction (Target 12) and invasive species' spread (Target 9).

As human societies increasingly utilize, restore and re-engineer natural systems, more species are affected by captive breeding, stocking and/or harvest. Simulations can determine probable consequences of these actions. Appropriate managed population sizes have been determined with simulations for a variety of organisms. Predictive simulations are also used to explore genetic and demographic impacts of harvest, for example revealing that selective logging may leave little negative, and even some beneficial, impact on some tropical tree species (Degen *et al.* 2006) – low impact logging may sometimes produce less inbreeding than nonlogging. More generally, simulations can help determine the optimal spatial distribution and size class for harvest. Lastly, simulations could forecast likely outcomes of hatchery-wild, crop-wild or native-introduced interactions in realistic landscapes (Perrier *et al.* 2012); predicting the speed and extent of hybrid replacement has been performed analytically but as yet individual-based simulations are not widely used.

An important question for molecular ecologists is whether to 'go genomic', or, more broadly, how many and which markers are needed for a given species and investigation goal (Andrew *et al.* 2013), a question that simulations can answer (Hoban *et al.* 2013a). For example, simulations were used to determine power of >700 SNP markers for estimating relatedness (Santure *et al.* 2010) in zebra finch, *Taeniopygia guttata*. However, even with hundreds of markers, relatedness estimates had low precision. Furthermore, due to linkage, more markers actually led to poorer estimates. Combining SNPs and microsatellites was not deemed useful for this task, but for detecting population structure combining marker types may work well (Narum *et al.* 2008).

Ancient DNA (from tissue or fossils up to thousands of years old) provides a unique means to examine species' histories. Coalescent simulators like BAYESSSC (Anderson *et al.* 2005) and FASTSIMCOAL (Excoffier & Foll 2011) are vital for interpreting ancient DNA data to infer past population processes (fragmentation, population size changes) and infer the timing and degree of changes, which may reveal their cause (climate, invasions, human exploitation). For example, in saiga antelope (*Saiga tatarica*), evidence of a strong bottleneck was not related to dramatic population loss due to hunting in recent decades, but to earlier bottlenecks and fragmentation (>2000 years before present, ybp), possibly from climate-induced habitat change from tundra to forest (Campos *et al.* 2010). In northern fur seal (*Callorhinus ursinus*), Pinsky *et al.* (2010) demonstrated that a large, northern refuge and high migration rates helped buffer against loss due to intense hunting. For grey whale (*Eschrichtius robustus*), simulations and ancient DNA (Alter *et al.* 2012) suggest that pre-bottleneck size was close to 100,000, that the cause was likely commercial whaling (100 ybp) and that the population nadir was ~10,000 (higher than estimated from historical records). These results have implications for evaluating threats to species, designing protected areas and determining policy.

## Cautions, outlook, conclusions

Simulation software are increasingly useful and accessible. However, simulations are sometimes criticized for creating detailed, highly parameterized models; thus, precise but incorrect predictions may result (Levins 1966; Peck 2004). Therefore, users of simulation software should take several cautions. First, a more complex model is not inherently a better one. Indeed, the best model is the simplest one that answers the question at hand, so users should avoid overparameterizing their simulations. In a foundational treatise on the subject, Levins (1966) advised modellers to test and compare results from both simple and complex models. Second, users must remember that simulations ultimately help bring a better but still imperfect understanding of the world, such as estimating the relative strength of different interacting forces, producing a probability distribution for a parameter, ranking management options or approximating threshold boundaries. Results obtained assume that the simulation model perfectly represents the real world, which it cannot. Thus, perhaps especially for complex simulations, output should be interpreted with caution and in combination with other evidence and expertise. Third, a range of values should be explored for parameters for which knowledge is scarce, for example movement rates or number of offspring (sensitivity analysis). Moreover, stochasticity should be included in environmental and demographic processes (Beissinger 2002). Another tactic is to combine multiple data sources (Table 1).

Estoup *et al.* (2010) showed that utilizing historical and genetic data in spatial simulations provided more accurate inference than utilizing only one data type. Simulating genotyping error (e.g. incorrect genotype calls, missing data) may be another important aspect of simulation-based studies, especially for ancient or non-invasive samples (Coombs *et al.* 2010). In real data sets, errors can cause information loss and decreased accuracy of inferred parameters or predictions, while systematic error can cause bias. Thus, imprecise or biased results may occur in real data sets but not simulated ones, which could lead to overly optimistic appraisal of simulation results if error is not included in the simulation model. Currently, only one simulator incorporates genotyping and sampling errors (PEDAGOG), so this feature is a priority for software developers. A related issue for simulating ever-larger data sets is to appropriately model variation in mutation and recombination. For example, some simulation-based studies model mutation at a constant rate (e.g. $\mu = 0.0005$ for microsatellites), but more realistic parameterization is possible (Estoup *et al.* 2001). Lastly, further work is needed on a concern particular to forward-in-time simulations – how long simulations should run to reach 'baseline' or quasi-equilibrium conditions and how to define and detect such conditions.

Readers will likely wonder how to choose the right type of simulator for each task. While some conventions exist (coalescent software are more often used for inference and forward-in-time software for prediction), there are few strict rules for this decision. Indeed, the flexibility of current software means that a large number of simulators can model multiple processes such as bottlenecks, migration, with an arbitrary degree of complexity (Hoban *et al.* 2012a). Nonetheless, no software features full complexity in all aspects of landscape, life history, dispersal, spatial dynamics, genetic architecture and natural selection. Particular software typically specialize in one or two of these features. Moreover, software differ in speed, ease-of-use, command-line capabilities or computer operating system.

I emphasize that it is not possible to provide exact recommendations of needed features for each of the seven categories of investigation, because strict requirements do not exist. A given investigation may be focused on the patch scale, across landscapes or between large populations. Additionally, the investigator may have different amounts of input data (historical, geographical, density). Depending on these and other aspects, different simulation software or features would be needed. Each of the seven categories explained above could use simulation techniques of varying sophistication, at different scales, with different levels of biological and ecological detail. In order to choose software suited to one's needs, at the outset of a simulation project investigators should devote substantial effort determining how best to model their system (e.g. continuous or discreet populations, temporal scale) and what features are likely to be important to include in the simulation model (adaptation, mating biology, overlapping generations, spatially explicit locations). For example, an investigation of local spatial genetic structure (SGS) or habitat resistance may require spatial locations and realistic dispersal kernels, as well as perhaps movement rules. An investigation about local adaptation would likely require variation in habitat quality as well as (possibly complex) genetic architecture and natural selection. Forward-in-time simulation software with options for life history (age of reproduction, clutch size, mating strategy, dispersal) are typically needed to infer species' biology (Fig. 2). Lastly, coalescent software is often used for long timescales or large population sizes. After making a list of properties of the system as well as study requirements (number of replications needed, length of simulation, population sizes, number and type of markers), the investigator can consult the Genetic Simulation Resources website or software reviews (Hoban *et al.* 2012a).

Choice of software will also depend on the methodology the investigator plans to use. For example, inference studies (Fig. 2) may employ either a relatively simple but nonetheless sufficient approach for comparing observed and simulated data, or may utilize approximate Bayesian computation. More generally, the dimensionality and completeness of parameter space that the investigator wishes to cover will vary from study to study. For exploring a large parameter space, an investigator may require a command-line program, which can be run using batch files and can be distributed on a computer cluster (Hoban *et al.* 2012a). Working in such space will require investigators to master basic scripting and bioinformatics skills (Haddock & Dunn 2010), in order to create input files with many combinations of parameter values, run analyses and collect and summarize output.

Simulations will of course not be necessary in every molecular ecology study. Some species or situations will conform relatively well to the assumptions of Bayesian or likelihood-based methods for estimating parameters of population size and migration rates. Straightforward analyses of parentage or assignment tests employing markers with a high information content (Hoban *et al.* 2012b; Veale *et al.* 2013) are another example in which simulations may not contribute much. Nonetheless, simulations have a role in most investigations by complementing analytical approaches or by verifying the power of the sampling strategy (Fig. 1).

It is important to note that differences in simulation software may lead to inconsistent results among studies, due to different underlying models and assumptions. For example, differing results were obtained from three recent studies (Landguth *et al.* 2010; Blair *et al.* 2012; Lloyd *et al.* 2013) regarding the lag time between establishment of a barrier and detection with genetic methods. All were individual based, but offspring distribution, generation overlap and dispersal model differed, as did the analytic method to determine a genetic signal. As Balkenhol & Landguth (2011) emphasize, this is not a weakness of simulations, but it does emphasize the importance of explicitly recognizing and describing assumptions, especially when generalizing results, and cautions against over-reliance on a single simulation software. It is highly recommended to check some of the main results by applying a second simulator software. Further, when planning complex simulations, investigators are encouraged to build input files in stages. By first executing simple simulation scenarios with clearly anticipated results, one gains practice and also ensures that distinct modules (e.g. selection, migration, population size) are parameterized correctly before creating complex simulations whose outcomes may not be known a priori.

Software from the case studies and examples in this review are SPLATCHE, IBDSIM, CDPOP, NEMO, QUANTINEMO, BAYESIAN SSC, FASTSIMCOAL, KERNELPOP, BOTTLESIM, MS and ALADYN. Software details and comparisons are reviewed elsewhere (Hoban *et al.* 2012a). Most simulation software is tailored to animals (although see KERNELPOP and ALADYN). More simulators featuring characteristics of plants or insects (polyploidy, large numbers of offspring, strong spatial mating patterns, complex dispersal, gene movement by seed and pollen) are needed, especially due to increasing interest in local adaptation of plants, demographic and genetic management of trees and wild crop relatives, and studies of plant–insect interactions. Lastly, no simulator explicitly models both a captive/managed population and wild populations, with ongoing exchange between, as practised in some zoo animals (sensu golden lion tamarin).

As data complexity increases and new techniques arise, there will continue to be a crucial role for simulators to evaluate power and bias of statistical methods in real-world situations, to determine when and how they should be applied (Epperson *et al.* 2010; Andrew *et al.* 2013) and to compare methods to identify the most powerful. For example, recent simulations found that some methods for early detection of population decline perform better than others under realistic parameters (Antao *et al.* 2011). Simulation-based evaluation could be a routine aspect of molecular ecology studies: investigators can use simulations to test how a method is expected to perform for their species, situation and molecular markers; to demonstrate sufficient power; and/or to optimize sampling or monitoring before a study begins (Blair *et al.* 2012; Landguth *et al.* 2012; Hoban *et al.* 2013b). Importantly, software developers should provide extensive documentation, including explanation of underlying models and assumptions. Users, in turn, must describe their methods explicitly including all parameter values. Clear documentation and explanation helps other investigators repeat simulation-based experiments to verify results, in perpetuity. Similarly, as simulators implement new features, there is a need to archive old versions and user manuals (Jones *et al.* 2006); the NEMO website is exemplary (http://sourceforge.net/projects/nemo2/). Scripts for analysing simulated data, parameter input files and/or simulated data sets can be shared online (e.g. Dryad database).

In the near future, genetic simulations will be combined with ecological, disease, atmospheric or hydrodynamic and/or multispecies models (Kremer *et al.* 2012) and will incorporate quantitative traits and selection (Kuparinen *et al.* 2010; De Mita *et al.* 2013). Such models can link to social and economic (Haight *et al.* 2002) or public health models (Magori *et al.* 2009), can be applied to managing genetic resources of domesticated species and may inform emerging issues like assisted colonization. Recent advances in data compression (Ruths & Nakhleh 2013) and data reduction (Aberer & Stamatakis 2013) hold particular promise for forward simulation of large sequences (>$10^6$ base pairs) for large numbers of individuals ($N > 10^5$), which are expensive for time and memory. The modular nature of new simulators also makes them easily expandable with new features (Landguth & Cushman 2010; Rebaudo *et al.* 2013; Strand & Niehaus 2007). As simplified models of the world, simulators lack perfect realism, are prone to overparameterization and may miss aspects of real systems (Levins 1966; May 2004). Simulation design and interpreting results (Figs 1–5) should be undertaken with caution and in a collaborative manner, with landscape and population ecologists, computational biologists, statisticians, population geneticists and management professionals (see Supporting Information, Appendix S1). Collaboration helps ensure full understanding of model assumptions/limits and full integration of available knowledge from multiple disciplines, enabling a bright future for population simulations in molecular ecology.

## Acknowledgements

I thank several anonymous reviewers, Frédéric Guillaume, Oscar Gaggiotti, Giorgio Bertorelle, Peter Galbusera, Carles Vilà, Michael Bruford, Allan Strand and Gernot Segelbacher for suggestions and encouragement. I acknowledge funding by European Commission Grant FP7-ENV-2009-1-244250 (Knowledge Transfer and Uptake), which established the ConGRESS (Conservation Genetic Resources for Effective Species Survival) project, as well as University of Ferrara, Italy. This project was completed while I was a Postdoctoral Fellow at the National Institute for Mathematical and Biological Synthesis (NIMBioS), an Institute sponsored by the US National Science Foundation, US Department of Homeland Security and US Department of Agriculture through NSF Awards #EF-0832858 and #DBI-1300426, with additional support from The University of Tennessee, Knoxville. I declare to my best knowledge no conflict of interest.

## References

Aberer AJ, Stamatakis A (2013) Rapid forward-in-time simulation at the chromosome and genome level. *BMC Bioinformatics*, **14**, 1–13.

Alberto F, Raimondi PT, Reed DC *et al.* (2010) Habitat continuity and geographic distance predict population genetic differentiation in giant kelp. *Ecology*, **91**, 49–56.

Alter SE, Newsome SD, Palumbi SR (2012) Pre-whaling genetic diversity and population ecology in Eastern Pacific gray whales: insights from ancient DNA and stable isotopes. *PLoS One*, **7**, 1–12.

Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics*, **21**, 1733–1734.

Andrew RL, Bernatchez L, Bonin A *et al.* (2013) A road map for molecular ecology. *Molecular Ecology*, **22**, 2605–2626.

Antao T, Pérez-Figueroa A, Luikart G (2011) Early detection of population declines: high power of genetic monitoring using effective population size estimators. *Evolutionary Applications*, **4**, 144–154.

Arenas M (2012) Simulation of molecular data under diverse evolutionary scenarios. *PLoS Computational Biology*, **8**, 1–8.

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

Banks SC, Ling SD, Johnson CR *et al.* (2010) Genetic structure of a recent climate change-driven range extension. *Molecular Ecology*, **19**, 2011–2024.

Beissinger SR (2002) Population Viability Analysis, past, present and future. In: *Population Viability Analysis* (eds Beissinger SR, McCullough DR), pp. 5–17. University of Chicago Press, Chicago, IL.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Blair C, Weigel DE, Balazik M *et al.* (2012) A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources*, **12**, 822–833.

Brekke P, Bennett PM, Santure AW, Ewen JG (2011) High genetic diversity in the remnant island population of hihi and the genetic consequences of re-introduction. *Molecular Ecology*, **20**, 29–45.

Bruford MW, Ancrenaz M, Chikhi L *et al.* (2010) Projecting genetic diversity and population viability for the fragmented orang-utan population in the Kinabatangan floodplain, Sabah, Malaysia. *Endangered Species Research*, **12**, 249–261.

Campos PF, Kristensen T, Orlando L *et al.* (2010) Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene. *Molecular Ecology*, **19**, 4863–4875.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues M, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, **323**, 785–789.

Carvajal-Rodriguez A (2010) Simulation of genes and genomes forward in time. *Current Genomics*, **11**, 58–61.

Caughley G (1994) Directions in conservation biology. *Journal of Animal Ecology*, **63**, 215–244.

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, **186**, 983–995.

Cook CN, Mascia MB, Schwartz MW, Possingham HP, Fuller RA (2013) Achieving Conservation Science that Bridges the Knowledge-Action Boundary. *Conservation Biology*, **27**, 669–678.

Coombs JA, Letcher BH, Nislow KH (2010) pedagog: software for simulating eco-evolutionary population dynamics. *Molecular Ecology Resources*, **10**, 558–563.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Currat M, Excoffier L (2011) Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *PNAS*, **108**, 15129–15134.

De Mita S, Thuillet A-C, Gay L *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, **22**, 1383–1399.

Degen B, Blanc L, Caron H *et al.* (2006) Impact of selective logging on genetic composition and demographic structure of four tropical tree species. *Biological Conservation*, **131**, 386–401.

Ejsmond MJ, Radwan J (2011) MHC diversity in bottlenecked populations: a simulation model. *Conservation Genetics*, **12**, 129–137.

England PR, Osler GHR, Woodworth LM *et al.* (2003) Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conservation Genetics*, **4**, 595–604.

Epperson BK, McRae BH, Scribner K *et al.* (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology*, **19**, 3549–3564.

Estoup A, Wilson IJ, Sullivan C, Cornuet J-M, Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics*, **159**, 1671–1687.

Estoup A, Baird SJE, Ray N *et al.* (2010) Combining genetic, historical and geographical data to reconstruct the dynamics of bioinvasions: application to the cane toad Bufo marinus. *Molecular Ecology Resources*, **10**, 886–901.

Excoffier L, Foll M (2011) fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, **27**, 1332–1334.

Ficetola GF, Bonin A, Miaud C (2008) Population genetics reveals origin and number of founders in a biological invasion. *Molecular Ecology*, **17**, 773–782.

Gapare WJ, Yanchuk AD, Aitken SN (2008) Optimal sampling strategies for capture of genetic diversity differ between core and peripheral populations of Picea sitchensis (Bong.) Carr. *Conservation Genetics*, **9**, 411–418.

Godefroid S, Piazza C, Rossi G *et al.* (2011) How successful are plant species reintroductions? *Biological Conservation*, **144**, 672–682.

Gossmann TI, Song B-H, Windsor AJ *et al.* (2010) Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular Biology and Evolution*, **27**, 1822–1832.

Guillaume F, Rougemont J (2006) Nemo: an evolutionary and population genetics programming framework. *Bioinformatics*, **22**, 2556–2557.

Haddock S, Dunn C (2010) *Practical Computing for Biologists*. Sinauer Associates, Sunderland, MA.

Haight RG, Cypher B, Kelly PA *et al.* (2002) Optimizing habitat protection using demographic models of population viability. *Conservation Biology*, **16**, 1386–1397.

Hailer F, Helander B, Folkestad AO *et al.* (2006) Bottlenecked but long-lived: high genetic diversity retained in white-tailed eagles upon recovery from population decline. *Biology Letters*, **2**, 316–319.

Hernandez RD (2008) A flexible forward simulator for populations subject to selection and demography. *Bioinformatics*, **24**, 2786–2787.

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

Hoban SM, Schlarbaum SE, Brosi SL, Romero-Severson J (2012b) A rare case of natural regeneration in butternut, a threatened forest tree, is parent and space limited. *Conservation Genetics*, **13**, 1447–1457.

Hoban SM, Gaggiotti OE, Bertorelle G (2013a) The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulation-based study. *Molecular Ecology*, **22**, 3444–3450.

Hoban SM, Gaggiotti OE, Bertorelle G (2013b) Sample Planning Optimization Tool for conservation and population Genetics (SPOTG): a software for choosing the appropriate number of markers and samples. *Methods in Ecology and Evolution*, **4**, 299–303.

Hoban S, Hauffe H, Perez-Espona S *et al.* (2013c) Bringing genetic diversity to the forefront of conservation policy and management. *Conservation Genetics Resources*, **5**, 593–598.

Hoban SM, Mezzavilla M, Gaggiotti OE *et al.* (2013d) High variance in reproductive success generates a false signature of a genetic bottleneck in populations of constant size: a simulation study. *BMC Bioinformatics*, **14**, 309.

Hudson RR (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Hundertmark KJ, Van Daele LJ (2010) Founder effect and bottleneck signatures in an introduced, insular population of elk. *Conservation Genetics*, **11**, 139–147.

Jones MB, Schildhauer MP, Reichman OJ, Bowers S (2006) The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 519–544.

Keinan A, Clark AG (2012) Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, **336**, 740–743.

Kimura M, Ota T, Ohta T (1975) Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proceedings of the National Academy of Sciences of the United States of America*, **72**, 2761–2764.

Kramer K, Degen B, Buschbom J *et al.* (2010) Modelling exploration of the future of European beech (*Fagus sylvatica* L.) under climate change—Range, abundance, genetic diversity and adaptive response. *Forest Ecology and Management*, **259**, 2213–2222.

Kremer A, Ronce O, Robledo-Arnuncio JJ *et al.* (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, **15**, 378–392.

Kuo C-HH, Janzen FJ (2003) bottlesim: a bottleneck simulation program for long-lived species with overlapping generations. *Molecular Ecology Notes*, **3**, 669–673.

Kuparinen A, Savolainen O, Schurr FM (2010) Increased mortality can promote evolutionary adaptation of forest trees to climate change. *Forest Ecology and Management*, **259**, 1003–1008.

Landguth EL, Cushman SA (2010) cdpop: a spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, **10**, 156–161.

Landguth EL, Cushman SS, Schwartz MK *et al.* (2010) Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, **19**, 4179–4191.

Landguth EL, Fedy BC, Oyler-McCance SJ *et al.* (2012) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Molecular Ecology Resources*, **12**, 276–284.

Leblois R, Estoup A, Rousset F (2009) IBDSim: a computer program to simulate genotypic data under isolation by distance. *Molecular Ecology Resources*, **9**, 107–109.

Levins R (1966) The strategy of model building in population biology. *American Scientist*, **54**, 421–431.

Lloyd MW, Campbell L, Neel MC (2013) The power to detect recent fragmentation events using genetic differentiation methods. *PLoS One*, **8**, e63981.

Lohmueller KE, Albrechtsen A, Li Y *et al.* (2011) Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genetics*, **7**, e1002326.

Magori K, Legros M, Puente ME *et al.* (2009) Skeeter Buster: a stochastic, spatially explicit modeling tool for studying Aedes aegypti population replacement and population suppression strategies. *PLoS Neglected Tropical Diseases*, **3**, 1–18.

Mardulyn P, Milinkovitch MC (2005) Inferring contemporary levels of gene flow and demographic history in a local population of the leaf beetle Gonioctena olivacea from mitochondrial DNA sequence variation. *Molecular Ecology*, **14**, 1641–1653.

Mardulyn P, Mikhailov YE, Pasteels JM (2009) Testing phylogeographic hypotheses in a euro-siberian cold-adapted leaf beetle with coalescent simulations. *Evolution*, **63**, 2717–2729.

Marino IAM, Benazzo A, Agostini C *et al.* (2013) Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation. *Molecular Ecology*, **22**, 5148–5161.

Marjoram P, Tavaré S (2006) Modern computational approaches for analysing molecular genetic variation data. *Nature Reviews Genetics*, **7**, 759–770.

Maruyama T, Kimura M (1980) Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *PNAS*, **77**, 6710–6714.

May RM (2004) Uses and abuses of mathematics in biology. *Science*, **303**, 790–793.

Messer PW (2013) SLiM: simulating evolution with selection and linkage. *Genetics*, **194**, 1037–1039.

Moran EV, Clark JS (2011) Estimating seed and pollen movement in a monoecious plant: a hierarchical Bayesian approach integrating genetic and ecological data. *Molecular Ecology*, **20**, 1248–1262.

Narum SRR, Banks M, Beacham TDTD *et al.* (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*, **17**, 3464–3477.

Naujokaitis-Lewis IR, Curtis JMR, Arcese P, Rosenfeld J (2009) Sensitivity analyses of spatial population viability analysis models for species at risk and habitat conservation planning. *Conservation Biology*, **23**, 225–229.

Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, **24**, 1552–1553.

Palmer SCF, Coulon A, Travis JMJ (2011) Introducing a "stochastic movement simulator" for estimating habitat connectivity. *Methods in Ecology and Evolution*, **2**, 258–268.

Paz-Vinas I, Quemere E, Chikhi L, Loot G, Blanchet S (2013) The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. *Molecular Ecology*, **22**, 3279–3291.

Peck SL (2004) Simulation as experiment: a philosophical reassessment for biological modeling. *Trends in Ecology & Evolution*, **19**, 530–534.

Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, **22**, 5972–5982.

Peng B, Chen H, Mechanic LE *et al.* (2013) Genetic Simulation Resources (GSR): a website for the registration and discovery of genetic data simulators. *Bioinformatics*, **29**, 1101–1102.

Pereira HM, Ferrier S, Walters M *et al.* (2013) Essential biodiversity variables. *Science*, **339**, 227–228.

Perrier C, Baglinière J-L, Evanno G (2012) Understanding admixture patterns in supplemented populations: combining molecular analyses and temporally explicit simulations in Atlantic salmon. *Evolutionary Applications*, **6**, 218–230.

Perrings C, Naeem S, Ahrestani F *et al.* (2010) Ecosystem services for 2020. *Science*, **330**, 323–324.

Petit RJ, Brewer S, Bordács S *et al.* (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, **156**, 49–74.

Pinsky ML, Newsome SD, Dickerson BR *et al.* (2010) Dispersal provided resilience to range collapse in a marine mammal: insights from the past to inform conservation biology. *Molecular Ecology*, **19**, 2418–2429.

Puebla O, Bermingham E, McMillan WO (2012) On the spatial scale of dispersal in coral reef fishes. *Molecular Ecology*, **21**, 5675–5688.

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.

Rebaudo F, Le Rouzic A, Dupas S *et al.* (2013) SimAdapt: an individual-based genetic model for simulating landscape management impacts on populations. *Methods in Ecology and Evolution*, **4**, 595–600.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation-by-distance. *Genetics*, **145**, 1219–1228.

Ruths T, Nakhleh L (2013) Boosting forward-time population genetic simulators through genotype compression. *BMC Bioinformatics*, **14**, 1–12.

Santure AW, Stapley J, Ball AD, Birkhead TIMR, Burke T, Slate J (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology*, **19**, 1439–1451.

Schiffers K, Bourne EC, Lavergne S, Thuiller W, Travis JMJ (2013) Limited evolutionary rescue of locally adapted populations facing climate change. *Philosophical Transactions of the Royal Society of London: B, Biological Sciences*, **368**, 20120083.

Schweitzer JA, Bailey JK, Rehill BJ *et al.* (2004) Genetically based trait in foundation tree affects ecosystem processes. *Ecology Letters*, **7**, 127–134.

Segelbacher G, Cushman SA, Epperson BK *et al.* (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, **11**, 375–385.

Shaw RG, Etterson JR (2012) Rapid climate change and the rate of adaptation: insight from experimental quantitative genetics. *New Phytologist*, **195**, 752–765.

Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, **19**, 3576–3591.

Stockwell CA, Hendry AP, Kinnison MT (2012) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, **18**, 94–101.

Strand AE, Niehaus JM (2007) kernelpop, a spatially explicit population genetic simulation engine. *Molecular Ecology Notes*, **7**, 969–973.

Travis JMJ, Munkemuller T, Burton OJ *et al.* (2007) Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and Evolution*, **24**, 2334–2343.

Veale AJ, Edge KA, McMurtrie P, Fewster RM, Clout MN, Gleeson DM (2013) Using genetic techniques to quantify reinvasion, survival and in situ breeding rates during control operations. *Molecular Ecology*, **22**, 5071–5083.

Wasserman TN, Cushman SA, Shirk AS, Landguth EL, Littell JS (2012) Simulating the effects of climate change on population connectivity of American marten (*Martes americana*) in the northern Rocky Mountains. *Landscape Ecology*, **27**, 211–225.

Wasserman TN, Cushman SA, Littell JS, Shirk AJ, Landguth EL (2013) Population connectivity and genetic diversity of

American marten (Martes americana) in the United States northern Rocky Mountains in a climate change context. *Conservation Genetics*, **14**, 529–541.

Whiteley AR, Coombs JA, Hudy M *et al.* (2012) Sampling strategies for estimating brook trout effective population size. *Conservation Genetics*, **13**, 625–637.

Yeaman S, Guillaume F (2009) Predicting adaptation under migration load: the role of genetic skew. *Evolution*, **63**, 2926–2938.

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution*, **65**, 1897–1911.

S.H. planned and wrote this article.

## Data accessibility

No data are associated with this article.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Table S1** Nearly 100 additional citations of simulation-based studies, divided into the seven categories of use and the particular task or interest of the investigators.

**Appendix S1** Additional guidance for employing population simulations.