Received: 13 November 2018 | Revised: 28 October 2019 | Accepted: 12 November 2019

DOI: 10.1111/1755-0998.13121

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Modelling multilocus selection in an individual-based, spatially-explicit landscape genetics framework

Erin L. Landguth[1] | Brenna R. Forester[2] | Andrew J. Eckert[3] | Andrew J. Shirk[4] | Mitra Menon[5] | Amy Whipple[6] | Casey C. Day[1] | Samuel A. Cushman[7]

[1]School of Public and Community Health Sciences, University of Montana, Missoula, MT, USA

[2]Department of Biology, Colorado State University, Fort Collins, CO, USA

[3]Department of Biology, Virginia Commonwealth University, Richmond, VA, USA

[4]Climate Impacts Group, College of the Environment, University of Washington, Seattle, WA, USA

[5]Integrative Life Sciences, Virginian Commonwealth University, Richmond, VA, USA

[6]Department of Biological Sciences and Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, AZ, USA

[7]USDA Forest Service, Rocky Mountain Research Station, Flagstaff, AZ, USA

Correspondence
Erin L. Landguth, School of Public and Community Health Sciences, University of Montana, Missoula, MT, USA.
Email: erin.landguth@mso.umt.edu

Funding information
National Science Foundation, Grant/Award Number: DEB-1340852, EF-1442486 and EF-1442597

Abstract

We implemented multilocus selection in a spatially-explicit, individual-based framework that enables multivariate environmental gradients to drive selection in many loci as a new module for the landscape genetics programs, CDPOP and CDMetaPOP. Our module simulates multilocus selection using a linear additive model, providing a flexible platform to evaluate a wide range of genotype-environment associations. Importantly, the module allows simulation of selection in any number of loci under the influence of any number of environmental variables. We validated the module with individual-based selection simulations under Wright-Fisher assumptions. We then evaluated results for simulations under a simple landscape selection model. Next, we simulated individual-based multilocus selection across a complex selection landscape with three loci linked to three different environmental variables. Finally, we demonstrated how the program can be used to simulate multilocus selection under varying selection strengths across different levels of gene flow in a landscape genetics framework. This new module provides a valuable addition to the study of landscape genetics, allowing for explicit evaluation of the contributions and interactions between gene flow and selection-driven processes across complex, multivariate environmental and landscape conditions.

KEYWORDS

CDMetaPOP, CDPOP, computer simulation, fitness surfaces, landscape resistance, natural selection

1 | INTRODUCTION

The programs CDPOP (Landguth & Cushman, 2010) and CDMetaPOP (Landguth, Holden, Mahalovich, & Cushman, 2017) are tools for individual-based simulation of mating, dispersal, and selection as functions of user-specified landscape resistance and selection processes (see Box 1 for a detailed comparison of the two programs). A module introduced in 2012 (Landguth, Cushman, & Johnson, 2012) enabled CDPOP to model natural selection based on spatial environmental fitness surfaces for one or two diallellic loci under directional selection (i.e., Gavrilets, 2004; Wright, 1932). That version of CDPOP required a spatial environmental fitness surface for each genotype in the one- or two-locus selection model. For example, in the single diallelic locus model, three relative fitness surfaces would be specified for the three genotypes (*AA*, *Aa*, and *aa*) from the two alleles, *A* and *a*, whereas nine surfaces would be required for the two-locus model. Selection in this module is implemented through differential survival of offspring as a function of the relative fitness of the offspring's genotype at the location where the dispersing individual settled.

While these simple models of natural selection have provided an important baseline for understanding adaptive landscape genetics, such as validation with theory (Landguth et al., 2012), comparing

methods for detecting genotype-environment associations (Forester et al., 2015; Jones et al., 2013), optimizing and informing conservation research (Creech et al., 2017; Landguth & Balkenhol, 2012; Landguth et al., 2017), and understanding the emergence of reproductive isolation in landscapes (Cushman & Landguth, 2016; Landguth, Johnson, & Cushman, 2015), the module cannot model more than two loci under selection. More flexible models linked to theory are needed to better represent complex genetic variation in real systems. For example, adaptive traits often have a complex genetic basis that interacts with selection strength, gene flow, drift, and mutation rate (Yeaman & Whitlock, 2011) in a multivariate environmental context; however, our ability to simulate these processes across many adaptive and neutral loci (e.g., hundreds to thousands) in a landscape genetic context has lagged.

Here, we incorporate multilocus selection into CDPOP using a linear additive model (Falconer & Mackay, 1996; Wade, Winther, & Goodnight, 2001), which enables the simulation of a broad range of genotype-environment associations. In the context of CDPOP, we define multilocus selection as multiple loci responding to univariate or multivariate environmental variables, impacting the local fitness of individuals with a given multilocus genotype. If selection is strong and is acting on a small set of loci with large effects, this could result in pronounced differences in allele frequencies across environments. However, if selection is distributed across a large number of loci, each with weak individual effects, multilocus selection will produce more subtle signals of allele frequency differentiation across environments (Le Corre & Kremer, 2003). These weaker signatures of selection, indicative of processes such as selection on standing genetic variation, are common patterns of natural selection in real populations (Bay et al., 2017; Le Corre & Kremer, 2012; Messer & Petrov, 2013; Pritchard & Di Rienzo, 2010) and are primarily what we sought to simulate in this new module.

In this article, we first introduce the framework for flexible, multivariate, multilocus simulation. We then validate the individual-based simulations under Wright-Fisher assumptions and evaluate expected results for simulations under a simple single landscape multilocus selection model. We then implement two examples of more realistic multilocus selection across complex landscapes with multiple loci and alleles linked to multiple environmental gradients. Our first example shows how multiple environmental variables can influence multilocus variation, while keeping the genotype space relatively simplistic (three loci). The second example shows how a suite of loci under different selection strengths can be simulated with varying gene flow in a landscape genetics framework, while keeping the environmental space relatively simplistic (one gradient environment).

2 | MATERIALS AND METHODS

2.1 | Simulation program

The spatial module of multilocus selection is built upon the existing framework of the individual-based landscape genetics program, CDPOP (Landguth & Cushman, 2010). CDPOP simulates genetic exchange and population dynamics for spatially referenced individuals on a resistance surface, where mating and dispersal events are a probabilistic function of effective or ecological distance between locations. As mentioned previously, past versions of CDPOP (e.g., Landguth et al., 2012) modelled natural selection via spatial environmental fitness surfaces for either one or two diallelic loci under directional selection (Gavrilets, 2004; Landguth et al., 2012; Wright, 1932), where the genotype-dependent viability of offspring was a function of their location on the fitness landscape. Here, we extend this approach to include options for simulating multilocus adaptive variation (i.e., more than two loci under selection) and multivariate environmental selection. Furthermore, this module can also be implemented within the branched program, CDMetaPOP (see Box 1; Landguth et al., 2017).

2.2 | Modelling multilocus selection association with environmental gradients

This new module incorporates multilocus selection from linear regression models (Falconer & Mackay, 1996; Wade et al., 2001), enabling the extension of landscape genomics analyses to explicitly and fully investigate adaptive evolution in complex landscapes. As with previous versions of CDPOP, the user specifies the genotype for each individual at the initial time step (i.e., number of loci and number of starting maximum alleles per locus). Now, the user also has the option of choosing any number of loci and alleles under selection, as well as any number of environmental variables that affect fitness for each allele. In this regression model, alleles at multiple loci associated with multiple environmental variables affect the local fitness (F^*) of an individual with a given multilocus genotype in an additive manner described by Equation 1:

$$F^* = b_0 + \sum_{i=1}^{n} \sum_{j=1}^{l} \sum_{k=1}^{a} X_i b_{ijk} A_{jk}. \quad (1)$$

The first term, b_0, provides the intercept of the linear model. The summation terms refer to the n environmental variables, and a number of alleles considered at l loci in an individual. Finally, for a given value of an environmental variable (X_i), we calculate the b_{ijk} effects of alleles A_{jk}, on fitness. Since statistical linkage will be generated in additive models whenever two or more loci experience simultaneous selection, we do not include a separate term for linkage disequilibrium here (Wade et al., 2001). Physical linkage is not included.

A fitness value, F, between 0 and 1 is obtained by rescaling Equation 1 by $(F^* - F^*_{min})/(F^*_{max} - F^*_{min})$, where F^*_{max} and F^*_{min} are the maximum and minimum, respectively, calculated from Equation 1 for all possible genotype-by-environment combinations. Rescaling the lowest fitness to 0 ensures there are no negative fitness values. F^*_{max} and F^*_{min} are calculated before simulations begin on the entire genotype-by-environmental space given the user defined b_{ijk} and X_i. Within the simulation workflow, CDPOP implements selection through differential survival $(1 - F)$ of an offspring given the relative fitness from Equation 1 at the

Box 1 Landscape genetic simulations in CDPOP and CDMetaPOP

Our landscape genetic simulation framework takes two current forms: CDPOP (Landguth & Cushman, 2010) and CDMetaPOP (Landguth et al., 2017). Both programs are individual-based, spatially-explicit, landscape demographic and genetic ('demogenetic') forward-in-time simulators. They simulate changes in neutral and/or selection-driven genotypes through time as a function of individual-based movement, complex spatial population dynamics, and multiple and changing landscape drivers modelled across continuous space. CDMetaPOP was a branch from CDPOP and initially developed for species living in dynamic riverine landscapes ('riverscapes'). However, CDMetaPOP can simulate a range of spatially-explicit processes that describe metapopulation theory across land, river, and seascapes, as well as accommodate simulations involving up to hundreds of thousands of individuals making it more computationally efficient than CDPOP. Both programs can be parameterized with empirical genetic, demographic, environmental, and/or spatial data sets.

Features in common between both CDPOP and CDMetaPOP include the capacity to simulate:

- Individual organisms and their growth, reproduction, genes, movement, mortality, etc. in a spatially-explicit environment.
- Spatially-explicit genetic processes such as gene flow, drift, mutation, bottlenecks, mtDNA, and neutral and adaptive genetics.
- Various demographic and reproductive processes, including philopatry, twinning, multiple paternity, asexual reproduction, differential subpopulation mortality, and age-structured mortality (e.g., Kristensen et al., 2018).
- Independent landscape surfaces for mating movement vs. dispersal: movement distributions are a result of cost distances (i.e., resistance) among locations.
- Dynamic landscape change at any user-defined spatial resolution, including changing land use or climate (Thatte, Joshi, Vaidyanathan, Landguth, & Ramakrishnan, 2018).
- Species reintroductions, translocations, augmentations, or invasions (Mims et al., 2019).
- Hybridization between two species, while tracking admixture through a hybridization index (Nathan et al., 2019).
- Disease dynamics, including vertical transmission of infection with a given probability and tracking of infected individuals.
- Local adaptation via genotype-environment association using a one- or two-locus selection model (Forester, Jones, Joost, Landguth, & Lasky, 2015; Jones et al., 2013; Landguth et al., 2012).
- Interactions between landscape resistance to gene flow and local adaptation (Scribner et al., 2016; Creech et al., 2017).

Features specific to CDMetaPOP include the capacity to simulate:

- Metapopulation dynamics and associated rates of dispersal, colonization, and extinction emerging from individual-based processes.
- Subpopulations where environmental conditions vary between, but not within, those subpopulations (i.e., patches).
- Up to hundreds of thousands of individuals.
- Class structure, specifically, size- or stage-based structure, that can control processes such as maturation probability, growth rate, migration and straying probabilities, fecundity, and capture probability (Day, Landguth, Bearlin, Holden, & Whiteley, 2018).
- Individual growth based on age, temperature, growing degree days, other environmental surfaces, or admixture coefficients.
- Selection at the genotype-environmental level or through genotype-maturation, genotype-growth, or genotype-stray/dispersal options.
- The hybrid index can also be used to implement differential fitness and/or assortative mating.
- Temporal systematic, stochastic, and demographic variability: processes and parameters can vary through time and can be linked spatially to environmental or climate variables at the patch level.
- Alternate mortality options, including density-independence, density-dependence at the patch and class level, logistic, and age/class-based. Multiple sources of mortality may operate in an additive or multiplicative fashion.
- Annual migration and straying behaviour based on cost-distance surfaces, including distinct demographic processes at migration grounds versus. natal grounds (Mims et al., 2019).
- Asymmetrical cost distance surface (e.g., wind surfaces: Landguth et al., 2017; riverine barrier surfaces: Mims et al., 2019).

This program note introduces a multilocus selection module that has now been implemented in both CDPOP and CDMetaPOP. This module uses a linear additive model to more flexibly parameterize selection across any number of loci and environmental or landscape variables. In combination with existing features, this new module will allow users to evaluate the contributions and interactions between gene flow, demography, and selection-driven processes across complex, multivariate environmental and landscape conditions.

CDPOP and CDMetaPOP are implemented with Python and require minimal scripting experience. Full details on installation and usage, including example runs and descriptions of input files and parameter settings, are provided in the CDPOP and CDMetaPOP user manuals, which are bundled with the program downloads.

location on the landscape where the dispersing individual settles. We provide a spreadsheet with CDPOP (betaFile_General.xlsx) that allows users to investigate the fitness impact of beta values in a simple two-locus, two-allele model with one or two environmental variables.

Users can also specify the genetic basis of local adaptation by modifying the environmental variables to reflect antagonistic pleiotropy (alternate alleles favoured in different environments) or conditional neutrality (alleles favoured in one environment but neutral in another; Anderson, Lee, Rushworth, Colautti, & Mitchell-Olds, 2013; Yoder & Tiffin, 2017). Because values for the environmental variables are spatially-explicit and can have very different scales of variability, we require that a standardization (z-score) is performed for each environmental variable (e.g., elevation, precipitation, land-use categories).

2.3 | Validation with global differential reproductive success

Validation of the new multilocus selection module required simulations of allele frequency change compared with the theoretical allele frequency change shown by Wright (1935) for both the single and double diallelic locus selection models. The theoretical equations and derivations can be found in Appendix 1 (i.e., Landguth et al., 2012) and are referred to herein as the Wright-Fisher model.

For the single diallelic locus selection model, we used Equation 1, $F = X_1([b_{111}A_{111} + b_{112}A_{112}])$, and set the average effects, b_{111} = 10 and b_{112} = −10. X_1 was a uniform spatial selection surface (i.e., Wright, 1935 assumption) with all values of 1. Thus, the genotypes *AA*, *Aa*, and *aa* had relative fitness values of 1.0, 0.5, and 0.0, respectively.

For the double diallelic locus selection model, we used Equation 1, $F = X_1([b_{111}A_{111} + b_{112}A_{112}] + [b_{121}A_{121} + b_{122}A_{122}])$, and set b_{111} = 10, b_{112} = −10, b_{121} = 10, and b_{122} = −10 with the uniform spatial selection surface, X_1, having all values of 1. Thus, the nine genotypes had relative fitness values of 1.0, 0.5, and 0.0 for groupings (*AABB*, *AABb*, *AaBB*), (*AAbb*, *AaBb*, *aaBB*), and (*Aabb*, *aaBb*, *aabb*), respectively.

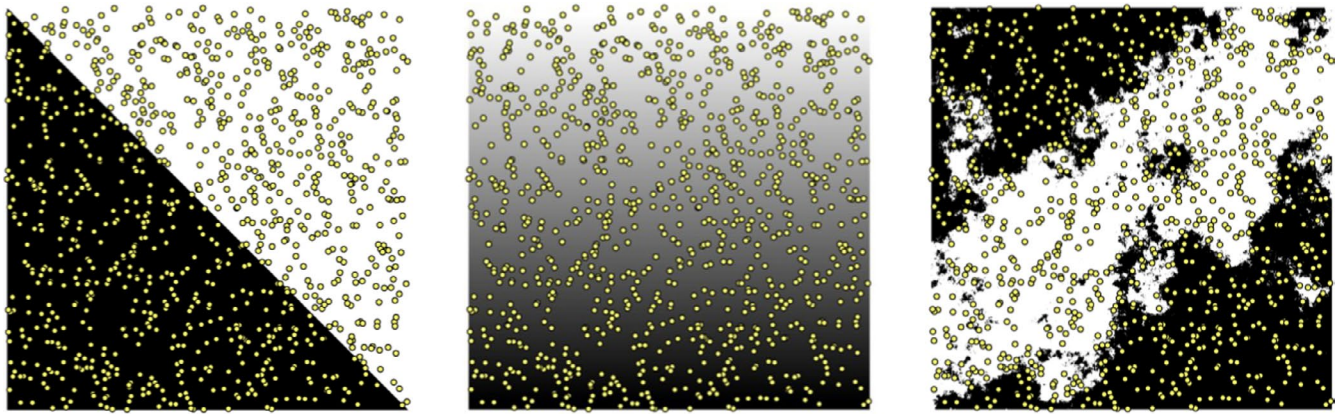
For both the single and double diallelic locus selection model simulations, the Wright-Fisher model was assumed (i.e., random mating, sexual reproduction with both female and male with replacement, offspring randomly disperse until a constant population is reached that has an equal sex ratio, no mutation, and nonoverlapping generations) with one exception: each mated pair produced two offspring to ensure a constant population size. We simulated individual genetic exchange across 100 nonoverlapping generations among 1,000 randomly spatially located individuals in a 1,024 × 1,024 gridded landscape for each selection model. All simulated populations contained an additional 50 diallelic neutral loci. The change in allele frequency for p_1, the allele frequency for *A*, was produced to compare with the results for the theoretical change in allele frequency via the Wright-Fisher selection models (Appendix 1). We ran 50 Monte Carlo replicates to assess variability in p_1 for each simulation.

2.4 | Expectations with spatially-variable differential reproductive success as a function of a single environmental variable

Our next task was to determine how spatially-variable selection implemented in the new simulation framework would affect spatial patterns of allele frequency. By removing a Wright-Fisher assumption, we stepped away from true theoretical validation as described in the previous section and moved into assessing whether the simulations met expectations. First, we conducted the single diallelic selection model as shown previously with the same simulation parameters, but replaced the uniform spatial selection surface with a spatially-variable selection surface. Using a 1,024 × 1,024 gridded raster, we created a categorical landscape that included an upper triangle with values of 1, a lower triangle with values of −1, and diagonal cells with a value of 0 (Figure 1a). Using Equation 1, $F = X_1([b_{111}A_{111} + b_{112}A_{112}])$, we now set the average effects, b_{111} = 10 and b_{112} = −10. Thus, individuals with the genotype *AA* would have a relative fitness values of 1.0 in the upper triangle area and 0.0 in the lower triangle area. Conversely, genotype *aa* would have relative fitness values of 0.0 in the upper triangle and 1.0 in the lower triangle. Individuals with the genotype *Aa* would have a relative fitness value of 0.5 regardless of where they settled on this landscape. To evaluate this spatial-selection simulation scenario, we ran three simulations which varied how we initialized the genotypes by starting the simulations with (i) only *AA*, (ii) only *aa*, and (iii) random assignment. At the end of the simulations, we plotted the distribution of each genotype (*AA*, *Aa*, and *aa*) on the landscape. We therefore would expect simulations to produce (i) only *AA* individuals in the upper triangle, (ii) only *aa* individuals in the lower triangle, and (iii) *Aa* individuals occurring anywhere on the landscape, with occurrences of *AA* and *aa* in only the upper and lower triangle, respectively.

2.5 | Expectations with spatially-variable differential reproductive success as a function of multiple environmental variables

Our next set of simulations included selection on multiple loci and alleles as a function of multiple environmental gradients, and with restricted isolation-by-distance dispersal. Here, we considered three environmental variables that affect fitness as shown in Figure 1, with three loci and two alleles per locus operating in the selection process. Our first environmental variable was the previously described categorical landscape (Figure 1a). The second environmental variable was a gradient landscape with continuous values ranging from 1 to −1 from the North-South (Figure 1b). The third environmental variable represented a fragmented landscape with equal proportion of values for one (e.g., favoured habitat) and −1 (e.g., nonfavoured habitat) created in the program QRULE (Gardner, 1999; H = 0.5 and p = .5; Figure 1c). Expanding Equation 1 to consider this example for three spatially variable selection



(a) X_1 – Categorical **(b)** X_2 – Gradient **(c)** X_3 – Habitat/No Habitat

FIGURE 1 Spatial selection landscapes with 1,000 simulated individuals. Values range from 1 to −1 represented by white to black, respectively. (a) Diagonal; (b) gradient and (c) habitat

gradients, three loci, and two alleles per locus requires 18 average effect sizes to be included (or three environmental variables * three loci * two alleles = 18 effect sizes), producing 27 possible genotypes (or $[(a + 2 - 1)!/(2! * (a - 1)!]^l$ where a is the possible alleles per locus [2 in this example] and l is the possible number of loci [three in this example]). For simplicity, we set b_{111} = 10 and b_{112} = −10 for the first locus and environmental variable, X_1, b_{221} = 10 and b_{222} = −10 for the second locus and environmental variable, X_2, and b_{331} = 10 and b_{332} = −10 for the third locus and environmental variable, X_3, where X_1, X_2, and X_3 are the diagonal, gradient, and habitat variables shown in Figure 1, respectively. The remaining betas were set to 0. We plotted the genotype-environment-fitness relationship for this spatially complex example in which three loci are associated with three environmental variables (Figure 2). For illustrative purposes and because we cannot visualize the full dimensional space, we held the first environmental variable constant at X_1 = 1 and third environmental variable constant at X_3 = 1. We allowed the second environmental variable to vary across its continuous space, X_2 = [−1, 1]. Equation 1 is calculated for each possible genotype by environment combination given the betas specified above and rescaled based on F^*_{max} and F^*_{min} (60 and −60, respectively) to achieve values between 0 and 1. Thus, using these simplified effect sizes, the genotypes that have the first allele present in each locus should be favored in areas that have values of 1 (i.e., upper triangle, towards north in the gradient landscape, and habitat patches). Unlike the previous panmictic movement simulations, we restricted movement of the individuals in these simulations to follow an inverse-square probability function constrained to a 25% maximum threshold of the entire landscape. We initialized all genotypes randomly at the start of the

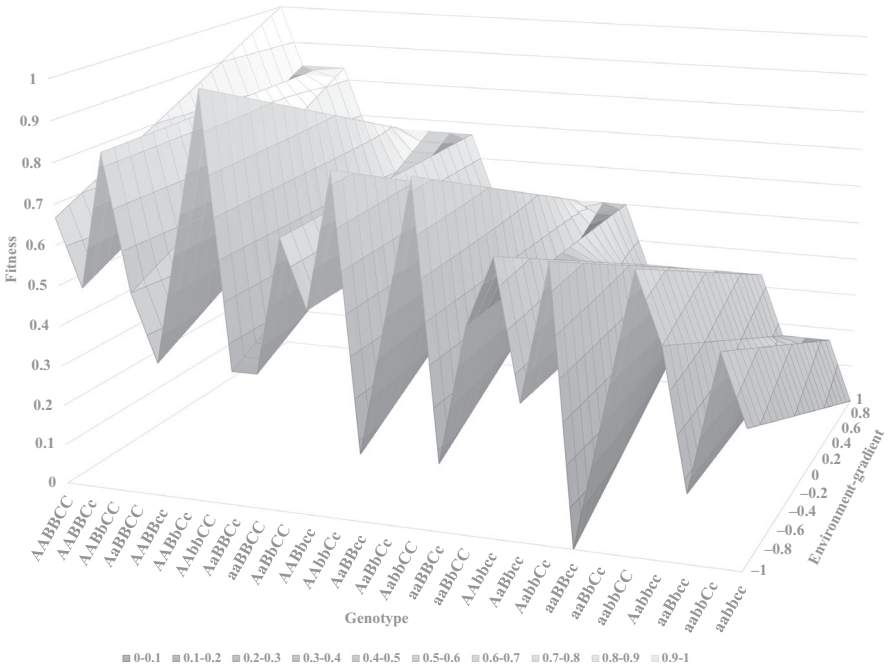


FIGURE 2 Genotype-environment-fitness landscape for the spatially complex example in which three loci are associated with three environments. For illustrated purposes, we held the first environmental variable constant at X_1 = 1 (diagonal landscape in Figure 1a) and third environmental variable constant at X_3 = 1 (habitat map in Figure 1c). We allowed the second environment to vary across its continuous space X_2 = [−1, 1] (gradient landscape in Figure 1b). Equation 1 is calculated for each possible genotype-by-environment combination given the betas specified (b_{111} = 10, b_{112} = −10, b_{221} = 10, b_{222} = −10, b_{331} = 10, b_{332} = −10, and 0 otherwise) and rescaled based on F^*_{max} = 60 and F^*_{min} = −60 to achieve fitness values between 0 and 1

simulations. To evaluate the results of this simulation, we plotted individuals on the landscape coded by their respective copies of the first allele at each of the three loci under selection. We therefore would expect to see individuals with homozygous copies of each locus to be more successful in landscapes patterned towards "1": upper triangle, towards the North, and in habitat patches.

2.6 | Expectations with multilocus selection and differential gene flow

Our final simulations illustrate how the new module can be used to simulate multilocus selection using a variety of selection strengths under different gene flow scenarios. Here, we simulated 1,000 loci with two alleles per locus: 100 loci under selection in response to a single environmental variable and 900 neutral loci. We used a gradient landscape with continuous values ranging from 1 to −1 from the North-South (Figure 1b). We set the first l = 1, 2, ..., 20 loci effect sizes to b_{111} = 0.15 and b_{112} = −0.15, the following l = 21, 22, ..., 50 loci effect sizes to b_{111} = 0.10 and b_{112} = −0.10, and the last l = 51, 52, ..., 100 loci effect sizes to b_{111} = 0.05 and b_{112} = −0.05 (reflecting "strong" [n = 20], "moderate" [n = 30], and "weak" [n = 50] selection, respectively). We increased the population size to 5,000 for these simulations within the same previous 1,024 × 1,024 simulation landscape. We restricted movement of the individuals in these final simulations to follow an inverse-square probability function constrained to 5%, 10%, and 15% maximum threshold of the entire landscape. We initialized all genotypes randomly at the start of the simulations and ran the simulations for 200 generations, using the first 100 generations as a burnin period where no selection was operating. To evaluate the results of these scenarios, we first used sparse non-negative matrix factorization to estimate the number of populations in each simulation (Frichot, Mathieu, Trouillon, Bouchard, & François, 2014) using the function *snmf* in the LEA package (Frichot & François, 2015) with R version 3.5.3 (R Core Team, 2019). We then calculated and plotted locus-specific F_{ST} for each of the 100 loci under selection and the first 400 neutral loci using the *wc* function in the HIERFSTAT package (Goudet & Jombart, 2015; note, for ease of visualization we did not plot all 900 neutral loci). Our expectation is for high, moderate, and weak differentiation of selected loci from neutral expectation for the strong, moderate, and weakly selected loci, respectively, with increasing levels of differentiation at both selected and neutral loci as gene flow decreases.

3 | RESULTS

3.1 | Multilocus selection under the Wright-Fisher model

For the single diallelic locus under the Wright-Fisher selection model, the CDPOP simulation nearly exactly matched Equation A2. $p_{1,t}$ (labelled Equation A2), and is plotted against the averaged simulated allele frequency for the 50 replicates in Figure 3a, despite the violation of one Wright-Fisher assumption (i.e., two offspring per mated pair). Thus, the new module implemented selection under "near" ideal Wright-Fisher conditions correctly and the outputs matched theoretical expectations. For the double diallelic locus scenario, the CDPOP simulation results also very closely matched the theoretical expectation of Equation A5. $p_{1,t}$ (labelled Equation A5) is plotted against the averaged simulated allele frequency for the 50 replicates in Figure 3b.

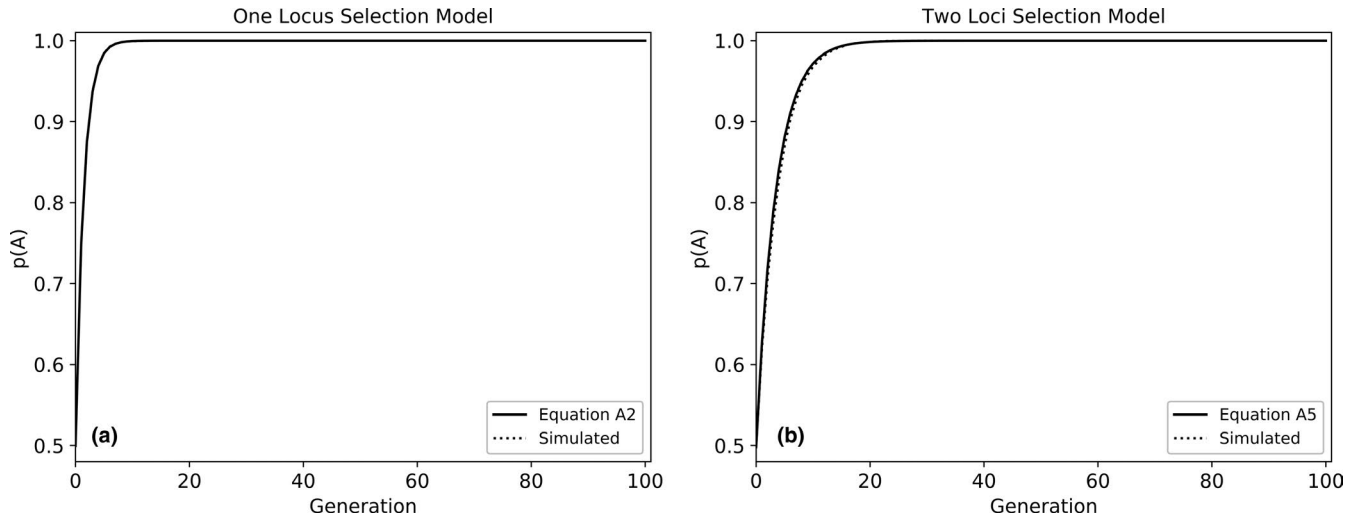


FIGURE 3 The simulations for the (a) single and (b) double diallelic locus selection models. For the single diallelic locus selection model, we used $F = X_1([b_{111}A_{111} + b_{112}A_{112}])$, and set the average effects, b_{111} = 10 and b_{112} = −10. For the double diallelic locus selection model, we used $F = X_1([b_{111}A_{111} + b_{112}A_{112}] + [b_{121}A_{121} + b_{122}A_{122}])$, and set set b_{111} = 10, b_{112} = −10, b_{121} = 10, and b_{122} = −10. X_1 was a uniform spatial selection surface with all values of 1. The dashed-dotted line is the averaged simulated allele frequency from the 50 replicates. The solid lines are the expected allele frequency given by Equations A2 and A5 (see Appendix 1). Note that the confidence intervals generated from the 50 Monte Carlo runs are too small to be viewed at this scale

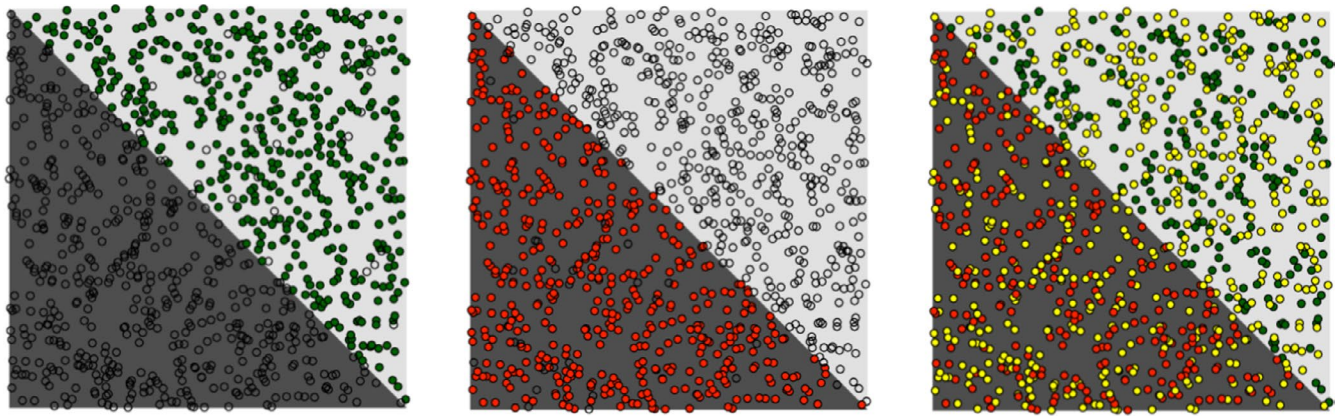


FIGURE 4 Outcome for simulation of a single spatially-variable selection landscape and single diallelic locus. Individuals are colour coded by genotypes: *AA*, green; *Aa*, yellow; *aa*, red. Empty circles indicate an unoccupied location. (a) Simulations were initialized with only *AA*. (b) simulations were initialized with only *aa*. (c) Simulations were initialized with randomly chosen genotypes

3.2 | Multilocus selection as a function of a single environmental variable

The next round of simulations enabled us to evaluate how the new module's spatially explicit functionality would affect resulting spatial patterns of allele frequency. Figure 4 shows the spatial plots for time 100 for the simulations with a single diallelic locus modelled on a categorical selection landscape (Figure 1a). The simulations that were only initialized with genotype *AA* showed that individuals could not survive within the lower triangle area (empty circles in Figure 4a). The opposite case is shown in Figure 4b, where individuals were initialized with only *aa*, resulting in survival only in the lower triangle. Random initialization of genotypes produced a spatial pattern consistent with antagonistic pleiotropy (i.e., different environmental conditions favor alternate alleles, Figure 4c). The homozygous genotypes emerged in their respective upper and lower triangles and genotype *Aa* (which had a relative fitness value of 0.5 regardless of spatial location) could occur anywhere on the landscape.

3.3 | Multilocus selection as a function of multiple environmental variables

Next, we illustrated a more complex scenario by simulating three loci, each under selection in response to a different environmental variable. Figure 5 shows individual genotypes for time step 100 plotted on the landscapes. For visualization, we superimposed each landscape shown in Figure 1 and white areas correspond to areas where each landscape had a value of 1 and black areas are areas where each landscape had a value of −1. For clarity, we coded individuals based on the count of the first allele at each of the three loci, where individuals could have either 2, 1, or 0 copies of this allele. Individuals with the most copies of these alleles (green dots) occupy areas in the upper triangle, towards the north for the gradient landscape, and in (grey) favoured habitat patches, as expected based on the spatial pattern of simulated fitness.

3.4 | Multilocus selection using varying levels of selection and gene flow

Our final simulations illustrated how the new module can simulate multilocus selection with loci under varying selection strengths. For each of the three dispersal scenarios (5%, 10%, 15%), we estimated five, four, and four populations, respectively, using the package *snmf*. After 100 generations of selection, we observed the expected signals of selection as a product of gene flow. Loci under strong selection (b_{111} = 0.15 and b_{112} = −0.15) were most differentiated from neutral expectations and displayed a positive correlation between variability in F_{ST} and gene flow (Figure 6). Likewise, loci under moderate ($b_{1/1}$ = 0.10 and $b_{1/2}$ = −0.10) and weak (b_{111} = 0.05 and b_{112} = −0.05) selection were less well differentiated, and displayed more variability in F_{ST} as dispersal capacity increased (Figure 6). Finally, we observed an inverse correlation between gene flow and population differentiation at both selected and neutral loci in response to increased isolation and genetic drift.

4 | DISCUSSION

We incorporated a multilocus selection module into the landscape genomic simulator CDPOP and its branched program, CDMetaPOP, using a linear additive modelling framework applied to multiple loci under selection as a function of multiple environmental variables. We showed that under Wright-Fisher assumptions, simulations of multilocus selection matched expected rates of changes in allele frequency. When spatially-explicit selection was implemented on environmental gradients, the new CDPOP module produced the expected spatial distribution of genotypes. This module allows for more realistic applications of multilocus selection across complex landscapes, with multiple loci and alleles under varying selection levels that can be linked to multiple environmental surfaces, providing an important tool for the rapidly growing field of landscape genomics.

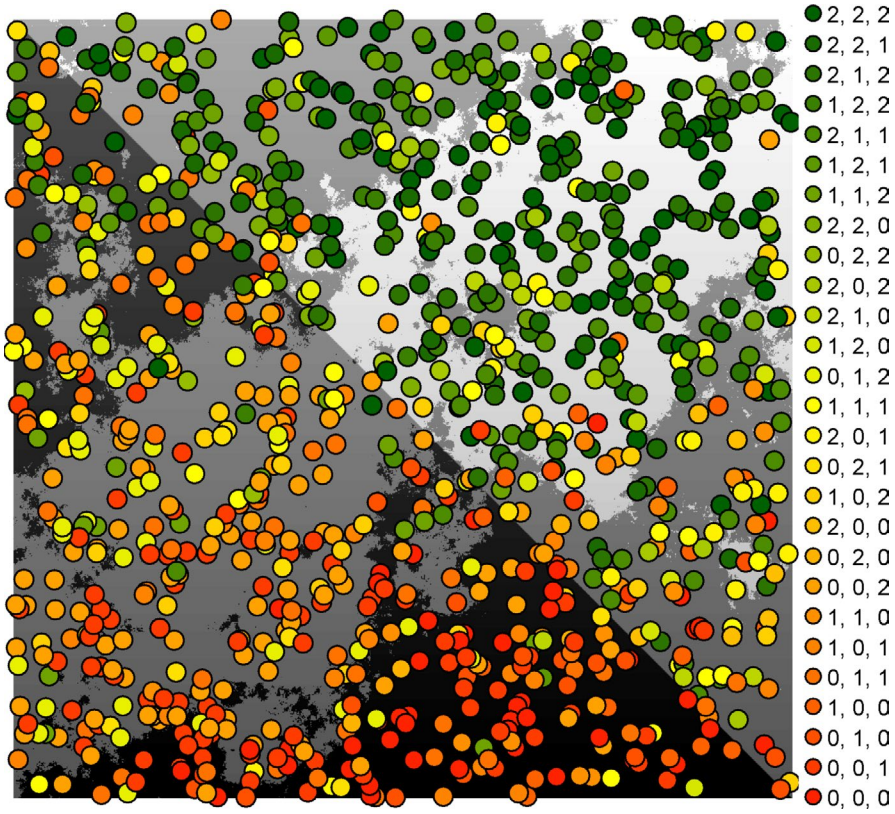


FIGURE 5 Outcome for simulation of a complex landscape and three loci. The three selection landscapes (Figure 1) are superimposed with lighter-white areas referring to areas where all three landscapes have values of 1 and darker areas mean all three landscapes have values of −1. The copies (either 2, 1, or 0) of the first allele for each of the three loci are plotted, where darker green genotypes have more copies of these alleles (e.g., 2, 2, 2 corresponds to 2 copies of the first allele for the first, second and third loci, respectively). The first locus is associated with the categorical landscape (X_1-Figure 1a). The second locus is associated with the gradient landscape (X_2-Figure 1b). The third locus is associated with the habitat fragmented landscape (X_3-Figure 1c)

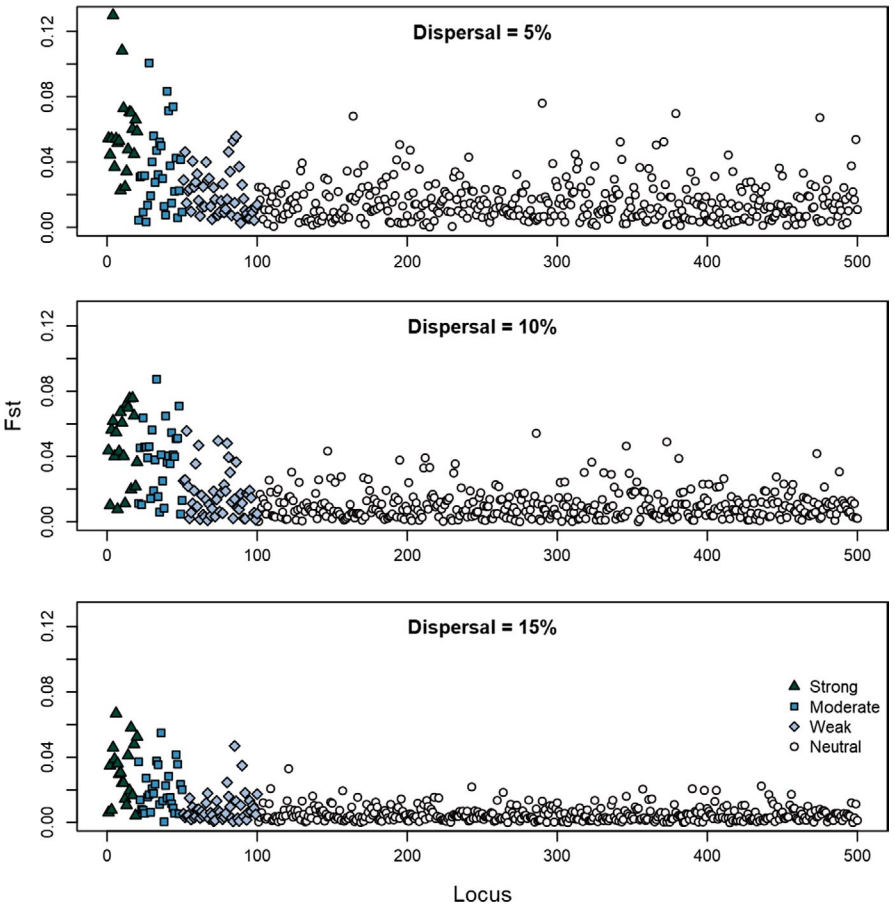


FIGURE 6 Locus-specific F_{ST} values for multilocus selection simulations under three selection strengths (strong, moderate, and weak) and three dispersal scenarios (5%, 10%, and 15% of the landscape). Note that the *y*-axis is on the same scale for the three simulations

The use of simulation modelling in landscape genetics is growing, as individual-based, spatially-explicit genetic modelling tools have proliferated (e.g., Nemo: Guillaume & Rougemont, 2006; quantiNemo: Neuenschwander, Michaud, & Goudet, 2018; SimAdapt: Rebaudo et al., 2013; MetaPopGen: Andrello & Manel, 2015; HexSim: Schumaker & Brookes, 2018). Landscape genetics studies are also increasingly coupling empirical data with simulation designs (e.g., Steinbach et al., 2018). Because essential model features and simplifying assumptions vary from study to study, the variety of available simulation models is a strength, since these programs focus on different processes and represent environmental heterogeneity in different ways. Landscape genetics researchers now have the opportunity to explore intermodel comparisons, providing further insight into the influences of model structure, parameters or other assumptions on landscape genetic processes (e.g., Safner, Miller, McRae, Fortin, & Manel, 2011; Blair et al., 2012). For a general review and comparison of available genetic simulation software, see Hoban (2014) or the Genetic Simulation Resources catalogue provided by the National Institute of Health (https://popmodels.cancercontrol.cancer.gov/gsr/).

Multilocus selection can be modelled in different ways. For example, SimAdapt (Rebaudo et al., 2013) implements a multiplicative model in which fitness values for single locus genotypes are multiplied to derive the fitness of the multilocus genotype. In this new module, we use a linear additive model of multilocus selection, which is no less biologically relevant than its multiplicative counterpart, and allows for linkage with additional relevant theory (Wade et al., 2001; Le Corre & Kremer, 2003; also see Appendix 1 for a more in-depth discussion of the use of additive vs. multiplicative models). No matter which model of selection is applied, landscape genetic simulators are limited by the computational burden of simulating many individuals (>10,000) with many loci (>20,000) across long time periods (>10,000 generations) and many Monte Carlo replicates (>1,000). In these computationally intensive simulation scenarios, other evolutionary population genomic programs may be more suitable (e.g., SLiM: Haller & Messe, 2017).

Interest in simulating realistic scenarios of multilocus selection, including the ability to model more complex multivariate and multilocus selection, has been driven by the growing number of empirical genomic data sets derived from next-generation sequencing (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016; Goodwin, McPherson, & McCombie, 2016). These technological and methodological advances have increased our capacity to investigate the genomic basis of local adaptation in an increasingly large number of non-model species. However, our ability to validate candidate adaptive loci that are identified through outlier-based approaches has lagged for a number of reasons, including: the small number of well-annotated reference genomes available; the difficulty of obtaining phenotypic data for many species (especially at the large sample sizes needed for robust genome-wide association studies to identify small effect loci that underlie most quantitative traits); and the impracticality of experimental approaches to validation for many organisms, such as common gardens and reciprocal transplants (e.g., Cushman, 2014; de Villemereuil, Gaggiotti, & Till-Bottraud, 2016; Lind, Menon, Bolte, Faske, & Eckert, 2018). In this context, simulation models may be the best option currently available for corroborating empirical findings in many species. Simulations can also be used to quantify error rates and conduct power analyses in landscape genomics studies, as they have proved useful in evaluating landscape genetic model selection approaches (e.g., Cushman & Landguth, 2010; Cushman, Wasserman, Landguth, & Shirk, 2013; Shirk, Landguth, & Cushman, 2018). We hope that this new module in CDPOP and CDMetaPOP will help researchers not only develop new theory in landscape genomics, but also improve the inference derived from empirical research.

ACKNOWLEDGEMENTS

This research was supported in part by funds provided by National Science Foundation grants EF-1442486, EF-1442597, and DEB-1340852.

AUTHOR CONTRIBUTIONS

The study was conceived by all. Funding was procured by E.L.L., S.C., A.J.E., A.W. Programming, simulations and analyses were completed by E.L.L., and B.F. All authors contributed to the manuscript and have approved the version for submission.

DATA AVAILABILITY STATEMENT

All data are available on DRYAD: https://doi.org/10.5061/dryad.d2547d7zp. CDPOP software and user manual are available on: https://github.com/ComputationalEcologyLab/CDPOP/. CDMETAPOP software and user manual are available on: https://github.com/ComputationalEcologyLab/CDMetaPOP/. All results can be recreated using CDPOP version 1.3.05.

ORCID

Erin L. Landguth https://orcid.org/0000-0002-7044-346X
Brenna R. Forester https://orcid.org/0000-0002-1608-1904
Andrew J. Eckert https://orcid.org/0000-0002-6522-2646
Andrew J. Shirk https://orcid.org/0000-0001-7700-5334

REFERENCES

Anderson, J. T., Lee, C.-R., Rushworth, C. A., Colautti, R. I., & Mitchell-Olds, T. (2013). Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, *22*, 699–708. https://doi.org/10.1111/j.1365-294X.2012.05522.x

Andrello, M., & Manel, S. (2015). MetaPopGen: An R package to simulate population genetics in large size metapopulations. *Molecular Ecology Resources*, *15*, 1153–1162.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*, 81–92. https://doi.org/10.1038/nrg.2015.28

Bay, R. A., Rose, N., Barrett, R., Bernatchez, L., Ghalambor, C. K., Lasky, J. R., … Ralph, P. (2017). Predicting responses to contemporary environmental change using evolutionary response architectures. *The American Naturalist*, *189*(5), 463–473. https://doi.org/10.1086/691233

Blair, C., Weigel, D. E., Balazik, M., Keeley, A. T. H., Walker, F. M., Landguth, E. L., … Balkenhol, N. (2012). A simulation-based evaluation of

methods for inferring linear barriers to gene flow. *Molecular Ecology Resources*, *12*, 822–833.

Creech, T. G., Epps, C. W., Landguth, E. L., Wehausen, J. D., Crowhurst, R. S., Holton, B., & Monello, R. J. (2017). Simulating the spread of selection-driven genotypes using landscaep reistance models for desert bighorn sheep. *PLOS ONE*, *12*(5), e0176960. https://doi.org/10.1371/journal.pone.0176960

Cushman, S. A. (2014). Grand challenges in evolutionary and population genetics: The importance of integrating epigenetics, genomics, modeling, and experimentation. *Frontiers in Genetics*, *5*, 197. https://doi.org/10.3389/fgene.2014.00197

Cushman, S. A., & Landguth, E. L. (2010). Spurious correlations and inferences in landscape genetics. *Molecular Ecology*, *19*, 3592–3602.

Cushman, S., & Landguth, E. L. (2016). Spatially heterogeneous environmental selection strengthens evolution of reproductively isolated populations in a Dobzhansky-Muller system of hybrid incompatibility. *Frontiers in Genetics*, *7*, 209. https://doi.org/10.3389/fgene.2016.00209

Cushman, S. A., Wasserman, T. N., Landguth, E. L., & Shirk, A. (2013). Re-evaluating casual modeling with Mantel tests in landscape genetics. *Diversity*, *5*, 51–72.

Day, C., Landguth, E. L., Bearlin, A., Holden, Z. A., & Whiteley, A. R. (2018). Using simulation modeling to inform management of invasive species: A case study of eastern brook trout suppression and eradication. *Biological Conservation*, *221*, 10–22. https://doi.org/10.1016/j.biocon.2018.01.017

de Villemereuil, P., Gaggiotti, O. E., Mouterde, M., & Till-Bottraud, I. (2016). Common garden experiments in the genomic era: New perspectives and opportunities. *Heredity*, *116*, 249–254. https://doi.org/10.1038/hdy.2015.93

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Harlow, UK: Longman Group.

Forester, B., Jones, M., Joost, S., Landguth, E. L., & Lasky, J. (2015). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, *25*, 104–120. https://doi.org/10.1111/mec.13476

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*, 925–929.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *194*, 973–983. https://doi.org/10.1534/genetics.113.160572

Gardner, R. H. (1999). QRULE: A program for the generation of random maps and the analysis of spatial patterns. In J. M. Klopatek, & R. H. Gardner (Eds.), *Landscape ecological analysis: Issues and applications* (pp. 280–303). New York, NY: Springer-Verlag.

Gavrilets, S. (2004). *Fitness landscapes and the origin of species*. Princeton, NJ: Princeton University Press.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*, 333–351. https://doi.org/10.1038/nrg.2016.49

Goudet, J., & Jombart, T. (2015). *hierfstat: Estimation and Tests of Hierarchical F-Statistics*. R package version 0.04-22.

Guillaume, F., & Rougemont, J. (2006). Nemo: An evolutionary and population genetics programming framework. *Bioinformatics*, *22*, 2556–2557. https://doi.org/10.1093/bioinformatics/btl415

Haller, B. C., & Messe, P. W. (2017). SLiM 2: Flexible, interactive forward genetic simulations. *Molecular Biology and Evolution*, *34*, 230–240. https://doi.org/10.1093/molbev/msw211

Hoban, S. (2014). An overview of the utility of popualtion simulation software in molecular ecology. *Molecular Ecology*, *23*, 2383–2401.

Jones, M. R., Forester, B. R., Teufel, A. I., Adams, R. V., Anstett, D. N., Goodrich, B. A., … Landguth, E. L. (2013). Integrating spatially-explicit approaches to detect adaptive loci in a landscape genomics context. *Evolution*, *67*, 3455–3468.

Kristensen, T. V., Puckett, E. E., Landguth, E. L., Hast, J., Carpenter, C., Sajecki, J. L., … Smith, K. G. (2018). Spatial genetic structure in American black bears (*Ursus americanus*): Female philopatry is variable and related population history. *Heredity*, *130*, 329–341.

Landguth, E. L., & Balkenhol, N. (2012). Relative sensitivity of neutral versus adaptive genetic data for assessing population differentiation. *Conservation Genetics*, *13*, 1421–1426.

Landguth, E. L., & Cushman, S. A. (2010). CDPOP: A spatially-explicit cost distance population genetics program. *Molecular Ecology Resources*, *10*, 156–161. https://doi.org/10.1111/j.1755-0998.2009.02719.x

Landguth, E. L., Cushman, S. A., & Johnson, N. A. (2012). Simulating natural selection in landscape genetics. *Molecular Ecology Resources*, *12*, 363–368. https://doi.org/10.1111/j.1755-0998.2011.03075.x

Landguth, E. L., Holden, Z. A., Mahalovich, M. F., & Cushman, S. A. (2017). Using landscape genetics simulations for planting blister rust resistance whitebark pine in the US Northern Rocky Mountains. *Frontiers in Genetics*, *8*, 9. https://doi.org/10.3389/fgene.2017.00009

Landguth, E. L., Johnson, N. J., & Cushman, S. A. (2015). Clusters of incompatible genotypes evolve with limited dispersal. *Frontiers in Genetics*, *6*, 151. https://doi.org/0.3389/fgene.2015.00151

Le Corre, V., & Kremer, A. (2003). Genetic variability at neutral markers, quantitative trait loci and trait in a subdivded population under selection. *Genetics*, *164*, 1205–1219.

Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, *21*, 1548–1566. https://doi.org/10.1111/j.1365-294X.2012.05479.x

Lind, B. M., Menon, M., Bolte, C. E., Faske, T. M., & Eckert, A. J. (2018). The genomics of local adaptation in trees: Are we out of the woods yet? *Tree Genetics and Genomes*, *14*, 29. https://doi.org/10.1007/s11295-017-1224-y

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, *28*, 659–669. https://doi.org/10.1016/j.tree.2013.08.003

Mims, M. C., Day, C. C., Burkhart, J. J., Fuller, M. R., Hinkle, J., Bearlin, A., … Landguth, E. E. (2019). Simulating demography, genetics, and spatially-explicit processes to inform reintroduction of a threatened char. *Ecosphere*, *10*, e02589. https://doi.org/10.1002/ecs2.2589

Nathan, L., Mamoozadeh, N., Tumas, H. R., Gunselman, S., Klass, K., Metcalfe, A., … Landguth, E. L. (2019). A simulation framework for evaluating fish hybridization dynamics in heterogeneous riverscapes. *Ecological Modeling*, *401*, 40–51.

Neuenschwander, S., Michaud, F., & Goudet, J. (2018). QantiNemo 2: A Swiss kife to simulate complex demographic and genetic scenarios, forward and backward in time. *Bioinformatics*, *35*, 886–888.

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation – Not by sweeps alone. *Nature Reviews Genetics*, *11*, 665–667. https://doi.org/10.1038/nrg2880

R Core Team, (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rebaudo, F., Rouzic, A. L., Dupas, S., Silvain, J.-F., Harry, M., & Dangles, O. (2013). SimAdapt: An individual-based genetic model for simulating landscape management impacts on populations. *Methods in Ecology and Evolution*, *4*, 595–600. https://doi.org/10.1111/2041-210X.12041

Safner, T., Miller, M. P., McRae, B. H., Fortin, M. J., & Manel, S. (2011). Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences*, *12*, 865–889.

Schumaker, N. H., & Brookes, A. (2018). HexSim: A modeling environment for ecology and conservation. *Landscape Ecology*, *33*, 197–211. https://doi.org/10.1007/s10980-017-0605-9

Scribner, K. T., Lowe, W. H., Landguth, E. L., Luikhart, G., Infante, D. M., Whelan, G., & Muhlfeld, C. C. (2016). Applications of genetic data

to improve management and conservation of river fishes and their habitat. *Fisheries*, *41*, 174–188.

Shirk, A., Landguth, E. L., & Cushman, S. A. (2018). A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular Ecology Resources*, *18*, 55–67. https://doi.org/10.1111/1755-0998.12709

Steinbach, P., Heddergott, M., Weigand, H., Weigand, A. M., Wilwert, E., Stubbe, M., … Frantz, A. C. (2018). Rare migrants suffice to maintain high genetic diversity in an introduced island population of roe deer (*Capreolus capreolus*): Evidence from molecular data and simulations. *Mammalian Biology*, *88*, 64–71. https://doi.org/10.1016/j.mambio.2017.11.009

Thatte, P., Joshi, A., Vaidyanathan, S., Landguth, E., & Ramakrishnan, U. (2018). Maintaining tiger connectivity and minimizing extinction into the next century: Insights from landscape genetics and spatially-explicit simulations. *Biological Conservation*, *218*, 181–191.

Wade, M. J., Winther, R. G., Agrawal, A. F., & Goodnight, C. J. (2001). Alternative definitions of epistasis: Dependence and interaction. *Trends in Ecology and Evolution*, *16*, 498–504. https://doi.org/10.1016/S0169-5347(01)02213-3

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress on Genetics*, *1*, 356–366.

Wright, S. (1935). Evolution in populations in approximate equilibrium. *Journal of Genetics*, *30*, 257–266. https://doi.org/10.1007/BF02982240

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

Yoder, J. B., & Tiffin, P. (2017). Effects of gene action, marker density, and time since selection on the perfomrance on landscape genomic scans of local adaptiation. *Journal of Heredity*, *109*, 16–28.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Landguth EL, Forester BR, Eckert AJ, et al. Modelling multilocus selection in an individual-based, spatially-explicit landscape genetics framework. *Mol Ecol Resour.* 2019;00:1–11. https://doi.org/10.1111/1755-0998.13121

MOLECULAR ECOLOGY RESOURCES

Supplemental Information for:

Modeling multilocus selection in an individual-based, spatially-explicit landscape genetics framework

Erin L. Landguth, Brenna R. Forester, Andrew J. Eckert,
Andrew J. Shirk, Mitra Menon, Amy Whipple, Casey C. Day, Samuel A. Cushman

Table of Contents:

MOLECULAR ECOLOGY
RESOURCES

Supplemental Text 1: Theoretical expected change in allele frequencies for one and two locus selection models.

The theoretical allele frequency change (Δp_1) as shown by Wright (1935) is

$$\Delta p_1 = \frac{p_1 q_1}{\bar{w}_1}[p_1(w_{11} - w_{12}) + q_1(w_{22} - w_{12})] \quad \text{[A1]}$$

where p_1 is the allele frequency for A, q_1 is the allele frequency for a, w_{11} is the relative fitness value for genotype AA, w_{12} is the relative fitness value for genotype Aa, w_{22} is the relative fitness value for genotype aa, and $\bar{w}_1 = p_1^2 w_{11} + 2p_1 q_1 w_{12} + q_1^2 w_{22}$ is the average fitness of the population. From here a difference equation for the single locus selection model can be derived to show the change in allele frequency for A through time,

$$p_{1,t} = \frac{p_{1,t-1}q_{1,t-1}w_{12} + p_{1,t-1}^2 w_{11}}{p_{1,t-1}w_{11} + 2p_{1,t-1}q_{1,t-1}w_{12} + q_{1,t-1}^2 w_{22}}. \quad \text{[A2]}$$

With the two-locus selection simulation, we compare the simulated change in allele frequency to the derived expected change in allele frequency. First, the expected frequency of the gametes after selection was

$$x'_{11} = \bar{w}_{11}(x_{11} - \mathrm{cD})/\,\bar{w}_2, \quad x'_{12} = \bar{w}_{12}(x_{12} + \mathrm{cD})/\,\bar{w}_2 \quad \text{[A3a,b]}$$

$$x'_{21} = \bar{w}_{21}(x_{21} + \mathrm{cD})/\,\bar{w}_2, \quad x'_{22} = \bar{w}_{22}(x_{22} - \mathrm{cD})/\,\bar{w}_2 \quad \text{[A3c,d]}$$

where the average relative fitness values for each gamete is $\bar{w}_{ij} = \sum_{k=1}^{2}\sum_{l=1}^{2} x_{kl} w_{ij.kl}$ and $\bar{w}_2 = \sum_{i=1}^{2}\sum_{j=1}^{2} x_{ij}\bar{w}_{ij}$ is the average relative fitness value for the population. We assumed independent assortment (i.e., free recombination, c = 0) and that there was no linkage disequilibrium (D = 0). Assuming $x_{11} = p_1 q_1$, $x_{12} = p_1 q_2$, $x_{21} = p_2 q_1$, and $x_{22} = p_2 q_2$, then the expected allele frequency after selection would be

$$p'_1 = x'_{11} + x'_{12}. \quad \text{[A4]}$$

Given the initial allele frequency of $p_1(0) = x_{11}(0) + x_{12}(0)$, the difference equation for the two locus selection model becomes

$$p_{1,t} = \frac{p_{1,t-1}q_{1,t-1}\bar{w}_{11} + p_{1,t-1}q_{2,t-1}\bar{w}_{12}}{\bar{w}_2}. \quad \text{[A5]}$$

MOLECULAR ECOLOGY
RESOURCES

Supplemental Text 2: Additive versus multiplicative models for fitness.

The convention of treating fitness as multiplicative across genotypes at a locus or across multiple loci is a historical assumption of the original models put forth by Fisher (1930) and Haldane (1932). As discussed in Wade et al. (2001), the reliance on the multiplicative model lies in mathematical convenience, rather than biological necessity. The argument below is only about multiplicative effects across genotypes and loci. Assuming multiplicative effects across life history stages during which selection manifests is a different issue (e.g., across mating types for evaluating fertility differentials, see Bodmer (1965).

For the single locus case with two alleles, this math (multiplicative fitnesses, e.g., AA: $(1+s)^2$, Aa: $1+s$, aa: 1) has the appealing property that the expected allele frequency change (Δp) across a single generation due to fitness differences is the same for both diploids and haploids (Felsenstein, 2017). This does not happen with additive effects on fitness (i.e., AA: $1+2s$, Aa: $1+s$, aa: 1). The relevant equations for the change in allele frequency across a single generation for each fitness scheme (add = additive, mul = multiplicative) are:

$$\Delta p_{add} = \frac{sp(1-p)}{1+2sp}$$

$$\Delta p_{mul} = \frac{sp(1-p)}{1+sp},$$

where p is the allele frequency of a reference allele in the current generation and s is the coefficient of selection. While the multiplicative model does allow for additional exact and easily derived solutions, this is solely a mathematical convenience and there is no biological reason to assume a multiplicative model for a single locus. When s is small, moreover, there is only a slight difference between these two assumptions (Figure A1) for Δp. Only at low to intermediate allele frequencies ($p \approx 0.2 - 0.5$) and very strong selection ($s > 0.8$) do the scaled expectations (i.e., divided by p) differ by approximately 10%. At more realistic values of s, the difference between assumptions is < 2%. Differences between these two parameterizations are

usually so small, moreover, that many authors assume they are equivalent (e.g., Smouse 1986) or have shown that the same qualitative patterns emerge (e.g., Karlin & Liberman, 1979a and references therein). In all cases, the assumption of additivity produces expected allele frequency changes between generations that are smaller than those assuming multiplicative effects. This introduces a conservative outcome to the assumption of the additive model, which may be preferred when using predications to inform management decisions.

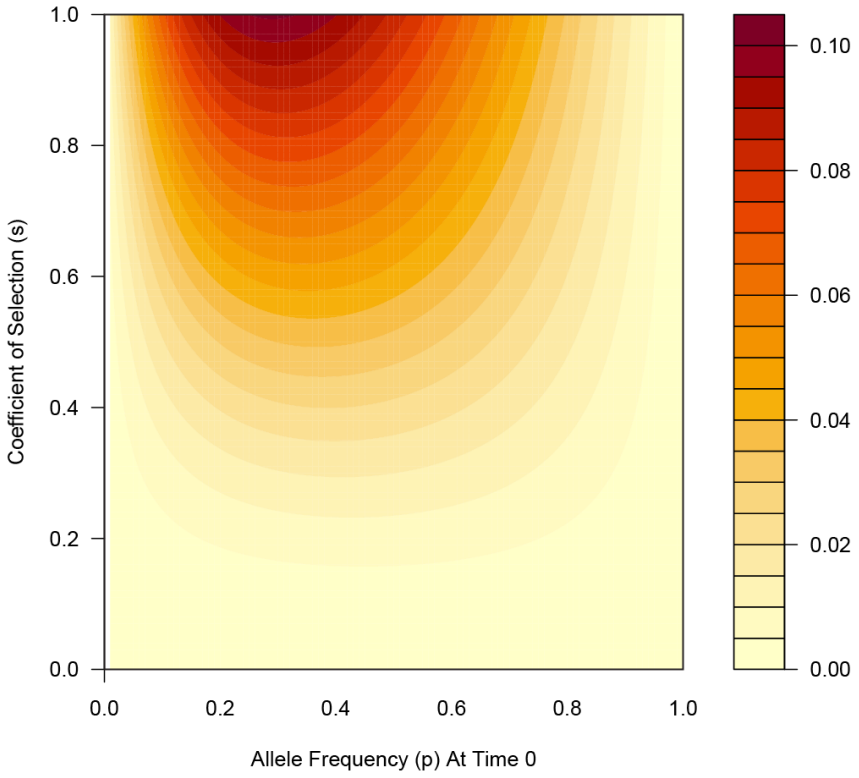


Figure A1. Relative difference in allele frequency change across a single generation for multiplicative and additive fitnesses as a function of allele frequency (p) and the coefficient of selection (s). Contours are based on the relative difference between multiplicative and additive assumptions, which is defined as the difference between Δp assuming multiplicative fitnesses and Δp assuming additive fitnesses divided by p.

For multiple loci, each with its own effects on viabilities and fecundities, multiplicative total fitness is calculated as the product of these locus-specific values. For example, if viability is the sole determinant of fitness, and two loci contribute to fitness differences across individuals, then the probability of survival for the multilocus genotype is the product of these probabilities for genotypes at each locus. Just as with additivity, however, this is an assumption, and one that enforces complete independence in the effects of alleles (and the genotypes they define for diploids) at each locus in determining overall fitness. Consider two-locus models of selection in a population that has a coefficient of linkage disequilibrium (D) equal to zero before selection. Assuming multiplicative fitnesses requires that D is also zero after selection, so that allele frequencies and their changes determine the response to selection. Thus, use of multiplicative fitnesses defines an architecture in this example with complete independence among its causative components (e.g., no linkage disequilibrium), so that much of their dynamics can be studied by simply considering the vector of allele frequency changes (one element per locus corresponding to the Δp for each locus, but see also Barton & Turelli 1987). This may be convenient mathematically, but does not necessarily provide increased adherence to biological reality relative to a completely additive parameterization.

Interestingly, a completely additive model across loci in two-locus selection models also has as its equilibrium a similar value of $D = 0$ (Karlin & Liberman 1979b), but this is an equilibrium that is approached over time, not an induced artifact of an assumption. The point here is to show, as does Wade et al. (2001), that there is nothing inherently more realistic biologically by assuming a multiplicative model, and in fact the additive model has additional relevant theory discussing dynamics of detectable signal in real data based on the build-up of linkage disequilibrium (e.g., Le Corre & Kremer 2003). Lastly, if we ignore linkage disequilibrium in either model, even strong selection (where most of the disconnect between models lies) on a sufficiently large multilocus architecture would distribute the total selective effect among so many loci that at any given locus this strong overall selection would manifest as a much smaller

coefficient of selection (Haldane 1930; Charlesworth & Charlesworth 2012, equation 3.17). We would thus be back to something like the case above (Figure A1) where the differences are relatively small between these models.

References:

Barton N, Turelli M (1987) Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genet. Res*., 49, 157-173.

Bodmer WF (1965) Differential fertility in population genetics models. *Genetics*, 51, 411–424.

Charlesworth B, Charlesworth D (2012) Elements of evolutionary genetics, Roberts and Company Publishers, Greenwood Village, CO (USA)

Felsenstein J (2017) Theoretical population genetics, University of Washington, Seattle, WA (USA) available at: http://evolution.gs.washington.edu/gs562/2017/pgbook.pdf

Fisher RA (1930) The genetical theory of natural selection, Oxford University Press, Oxford (UK).

Haldane JBS (1930) A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. *Proc. Camb. Philos. Soc*., 27, 131-136.

Haldane JBS (1932) The causes of evolution. Macmillan, Oxford (UK).

Karlin S, Liberman U (1979a) Central equilibria in multilocus systems. I. Generalized nonepistatic selection regimes. *Genetics*, 91, 777-798.

Karlin S, Liberman U (1979b) Representation of nonepistatic selection models and analysis of multilocus Hardy-Weinberg equilibrium configurations. *J. Math. Biol*., 7, 353–374.

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics*, 164, 1205-1219.

Smouse P (1986) The fitness consequences of multiple-locus heterozygosity under the multiplicative overdominance and inbreeding depression models. *Evolution*, 40, 946-957.

Wade MJ, Winther RG, Agrawal AF, Goodnight CJ (2001) Alternative definitions of epistasis: dependence and interaction. *TREE*, 16, 498-504.

Wright S (1935) Evolution in populations in approximate equilibrium. *J. Genetics*, 30, 257–266.