TECHNICAL NOTE

CDFISH: an individual-based, spatially-explicit, landscape genetics simulator for aquatic species in complex riverscapes

Erin L. Landguth · C. C. Muhlfeld ·
G. Luikart

Abstract We introduce Cost Distance FISHeries (CDFISH), a simulator of population genetics and connectivity in complex riverscapes for a wide range of environmental scenarios of aquatic organisms. The spatially-explicit program implements individual-based genetic modeling with Mendelian inheritance and k-allele mutation on a riverscape with resistance to movement. The program simulates individuals in subpopulations through time employing user-defined functions of individual migration, reproduction, mortality, and dispersal through straying on a continuous resistance surface.

Keywords Stream resistance · Conservation biology · Freeware · Gene flow · Computer simulations · Stream networks · Dispersal · Connectivity · Habitat fragmentation · Climate change

The importance of simulation modeling in landscape genetics has been emphasized in a number of recent articles

E. L. Landguth (✉)
Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA
e-mail: erin.landguth@mso.umt.edu

C. C. Muhlfeld
U.S. Geological Survey, Northern Rocky Mountain Science Center, Glacier National Park, West Glacier, MT 59936, USA

G. Luikart
Division of Biological Sciences, Flathead Lake Biological Station, University of Montana, Polson, MT 59860, USA

(e.g., Balkenhol et al. 2009; Epperson et al. 2010; Balkenhol and Landguth 2011). However, available computational tools for modeling complex stream networks (riverscapes) are limited. We developed CDFISH to model gene flow in riverscapes from the program architecture of CDPOP (Landguth and Cushman 2010). There are three major new functionalities in CDFISH that are not included in CDPOP or any other simulation program. First, individuals are spatially-explicit and designated in subpopulations. Second, subpopulations are allowed to reach a carrying capacity, stabilize, or become extinct. Third, dispersal back to subpopulations is modeled through residency, migration, and straying.

These CDFISH features enable genetic simulations of a wider range of biological scenarios than available in other programs (e.g., AQUASPLATCHE (Neuenschwander 2006), GENE-NET (http://www.inh.fr/pageperso/cfleuran/genenet/genenet.html)). CDFISH differs in several ways, which (1) is individual-based rather than deme-based, (2) is a forward simulator for genotypes of all individuals in a spatially structured landscape, (3) dispersal and mating are explicitly cost functions across heterogeneous riverscapes. CDFISH is specifically designed to enable quantification of how stream resistance (to movement) affects gene flow patterns and the time required for spatial patterns of genetic relatedness to change or equilibrate as functions of individual-based movement (mating and dispersal), vital rate dynamics, and mutation on flexible resistance landscapes.

The program is written in Python 2.6 and is provided with instructions for most platforms, along with sample files. CDFISH is built on a docking architecture that allows for ease of future development. CDFISH has been debugged by testing all combinations of options. A user manual, example data sets, and free downloads can be found at http://cel.dbs.umt.edu/software/CDFISH/. Here, we briefly

describe the main features noting that a more detailed description can be found in the user manual.

CDFISH models genetic exchange for a stream resistance surface and (x, y) located individuals as functions of individual-based movement (mating and dispersal), vital dynamics, and mutation. Users must create a stream resistance surface where each cell value (pixel) represents the unit cost of crossing each location. Pixels are given weights or 'resistance values' reflecting the presumed influence of each variable to movement or connectivity of the species in question (e.g., Dunning et al. 1992; Cushman et al. 2006; Spear et al. 2010). Stream resistance surfaces can be parameterized to reflect different costs to movement associated with water temperature and flow, habitat complexity, physical barriers, elevation, slope, or other stream features. From here, an input matrix of movement costs can be computed between all pairs of individuals (e.g., by using tools such as UNICOR (Landguth et al. in press) or CIR-CUITSCAPE (McRae and Beier 2007)). By comparing genetic distances between individuals with ecological cost distances between them, researchers can test specific hypotheses about the influences of stream features and

environmental conditions on gene flow (Cushman et al. 2006; Epps et al. 2007). The program can flexibly incorporate absolute or partial barriers and panmixia within this framework.

Individual locations are user defined within subpopulations. The genotypes are initialized randomly or with empirical data. The initial age structure and sex of each individual is user defined. Reproduction is heterosexual with a random mating structure. Mated pairs are chosen based on with or without replacement combinations. Each mated pair can have a number of offspring that is a bounded random draw, a specified mean Poisson draw, or a constant number. Mendelian inheritance with k-allele mutation is used to generate the genotypes.

Offspring can either reside in each subpopulation (e.g., tributary) or disperse to a source location (e.g., lake, river, or ocean). Offspring that disperse to the source must migrate back to their respective subpopulation or stray. Straying and residency probabilities are user defined. Migration is a function of the cost to movement in the resistant riverscape. To reflect species dispersal abilities, the user can specify the maximum effective distance an

Fig. 1 An example riverscape with 19 subpopulations and 50 individuals per subpopulation. The *left inset* shows a zoomed in spawning bed with 50 individual locations. The *right inset* shows the zoomed in riverscape with varying degrees of resistance values assigned to each pixel value in the stream network. The "Lake Source" is Flathead Lake (Montana, USA), to which some species that have a migratory life history (e.g., *Salvelinus confluentus*) will migrate to
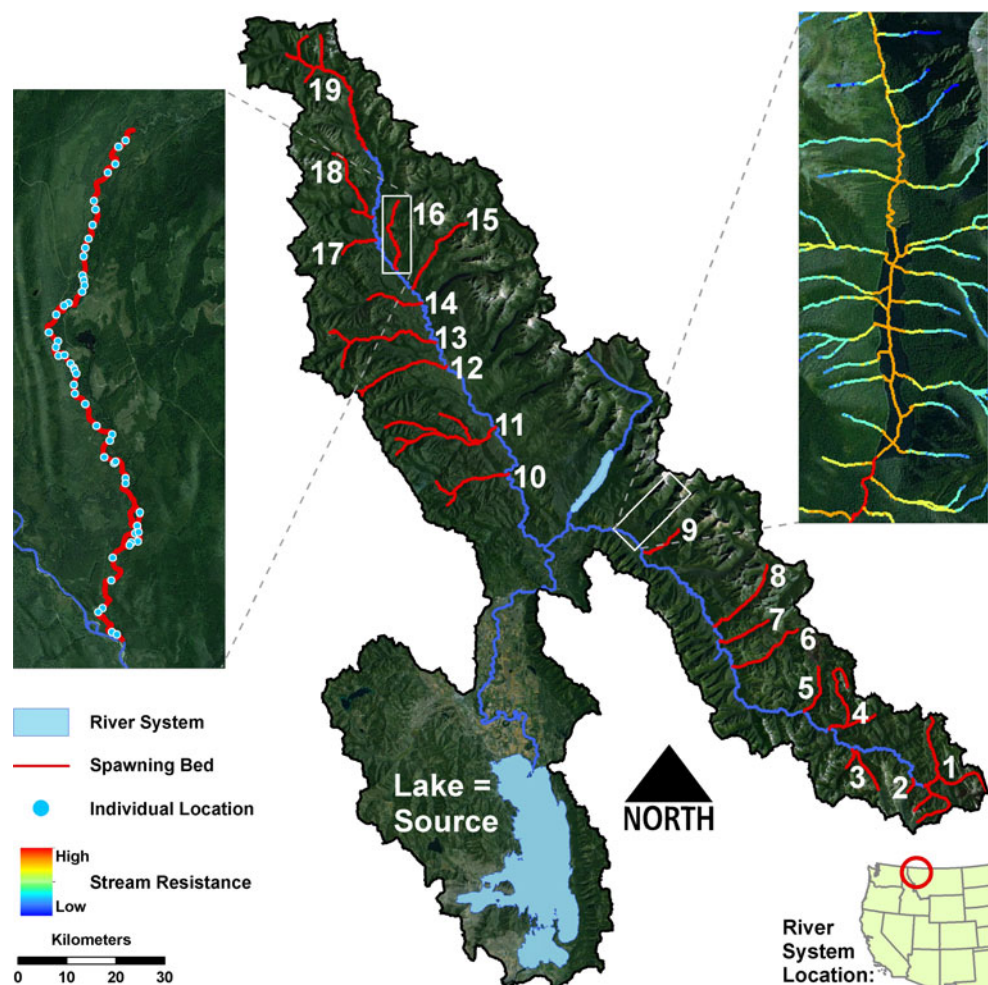
individual can travel in cost units. Offspring move a distance from their source location based on a random draw from a probability distribution inversely proportional to a user-specified function (e.g., linear).

The user must specify the parameters through an input script file (Table S1). As the model simulates stochastic processes, most applications will benefit from quantification of the mean and variability of genetic structure and Monte Carlo option is provided. Additionally, the user may also wish to conduct sensitivity analyses and is provided through batch capability. The program outputs files for the genotypes, age, and sex of each individual at specified generations, along with a file containing various population parameters.
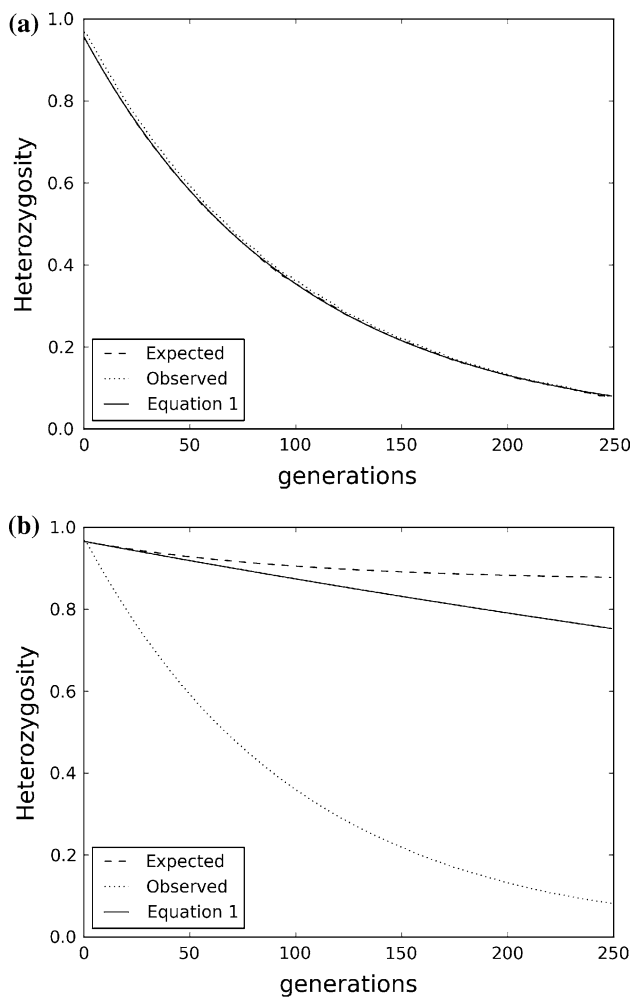

Fig. 2 CDFISH heterozygosity results from the two examples. **a** The overlapping simulations with Eq. 1 H_t show the theoretical rate of decay in the panmictic scenario, as expected for one subpopulation ($N = N_e = 50$). **b** The non-overlapping simulations with Eq. 1 H_t show Whalund effects as expected from population genetics whenever there is substructure (Allendorf and Luikart 2007), i.e., all subpopulations are pooled to calculate H_t, but H_e and H_o are computed from the mean among subpopulations

Comparisons of simulation output with theoretical equations illustrate the validation, usefulness, and reliability of CDFISH. Genetic exchange is simulated in 19 subpopulations with 50 individuals per subpopulation under a riverscape (Fig. 1). Wright-Fisher assumptions were used for 250 generations, 100 Monte Carlo replicates, and genotypes initialized with 30 loci and 30 alleles per locus. We provide basic graphics, including comparison of results to theoretical predictions for loss of heterozygosity over time (Fig. 2).

H_e and H_o are calculated at each generation as the estimated proportion of individuals that are expected heterozygous based on a Hardy–Weinberg population and the proportion of individuals that are observed heterozygous, respectively. H_e and H_o results are shown in Fig. 2, compared with heterozygosity produced according to the following equation (Crow and Kimura 1970) adjusted for a small sample size,

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

where the effective population size for separate sexes (Hedrick 2011) is defined as

$$N_e = \frac{4N_M N_F}{N_M + N_F},$$ (2)

and H_t is the theoretical rate of decay after t generations, $H_o(0)$ is initial heterozygosity, N_M is the male total, and N_F is the female total.

References

Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell, Malden

Balkenhol N, Landguth EL (2011) Simulation modeling in landscape genetics: on the need to go further. Mol Ecol 20:667–670

Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, Holderegger R, Wagner HH (2009) Identifying future research needs in landscape genetics: where to from here? Landscape Ecol 24:455–463

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York

Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with casual modeling. Am Nat 168:486–499

Dunning JB, Danielson BJ, Pulliam HR (1992) Ecological processes that affect populations in complex landscapes. OIKOS 65: 169–175

Epperson BK, McRae BH, Scribner K, Cushman SA, Rosenberg MS, Fortin M-J, James PMA, Murphy M, Manel S, Legendre P, Dale MRT (2010) Utility of computer simulations in landscape genetics. Mol Ecol 19:3549–3564

Epps CW, Wehausen JD, Bleich VC, Torres SG, Brashares JS (2007) Optimizing dispersal and corridor models using landscape genetics. J Appl Ecol 44:714–724. http://nature.berkeley.edu/BrasharesLab/documents/Epps%20et%20al%20J_Appl_Ecol%202007%20with%20Appendices.pdf

Hedrick PW (2011) Genetics of populations. Jones and Bartlett Publishers, Sudbury

Landguth EL, Cushman SA (2010) CDPOP: a spatially-explicit cost distance population genetics program. Mol Ecol Res 10:156–161

Landguth EL, Hand BK, Glassy JM, Cushman SA (2011) UNICOR: a species connectivity and corridor network simulator. Ecography (in press)

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Science USA 104:19885–19890

Neuenschwander S (2006) AQUASPLATCHE: a program to simulate genetic diversity in populations living in linear habitats. Mol Ecol Notes 6:583–585

Spear S, Balkenhol N, Fortin M-J, McRae B, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Mol Ecol 19:3576–3591