NEW APPROACHES IN LANDSCAPE GENETICS AND NICHE MODELING FOR UNDERSTANDING LIMITS TO ANURAN DISTRIBUTIONS

By

MELANIE APRIL MURPHY

A dissertation submitted in partial fulfillment of
The requirement for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Biological Sciences

August 2008

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of MELANIE APRIL MURPHY find it satisfactory and recommend that it be accepted.

_____
Chair

_____

_____

_____

ACKNOWLEDGMENT

First, I would like to thank my committee for their support and multiple contributions: Andrew Storfer. Ray Dezzani, Chuck Peterson, and Mike Webster. In addition, I would like to thank the following people for helpful comments and discussions: P Bartelt, S Cushman, J Evans, C Goldberg, J Kerby, R Klaver, M Morgan, D Patla, D Pilliod, A Robinson, S Spear, C Steele, and L Waits. A special thank you goes to the Palouse Landscape Genetics and the Storfer lab groups for feedback on ideas and manuscript drafts. I also thank One World Café for providing a space for many of these discussions and much of the writing of this dissertation.

I would like thank M Murphy, M Flisher, and D Murphy for field assistance in Yellowstone; C Corkan, D Corkan, D Patla, and C Peterson provided additional tissue samples. I would like thank S. Blaser, J. Boone, D. Esperas, E. Letterman, E. Suronen, M. E. Valasquez, and J. Visciglia for field assistance in the Bighorn Crags. In addition, I thank D Borcard, N Crookston, and D Hall for modifying computer code. A Gallant and J Shive assisted in compiling spatial data.

Research was permitted by Yellowstone National Park (Permit YELL-05452**;** C Smith and C Hendrix), Idaho Fish and Game (Wildlife Collecting Permit # 030716), and IACUC (ASAF #3378). Funding for this research was provided by: National Science Foundation award number DEB-0606458, Amphibian Research and Monitoring Initiative (ARMI), SBS King Fellowship, Washington State University Zoology Alumni Scholarship, EPA-STAR Fellowship (#FP-916695), Theodore Roosevelt Memorial Fund, Society for Wetland Science (2004, 2005), Canon National Parks, Sigma-Xi grants-in-aid (2004), and Graduate Women in Science.

Finally, I thank Zoe, Mariah, Dru, Widgy and Menika for keeping it all in perspective.

NEW APPROACHES IN LANDSCAPE GENETICS AND NICHE MODELING FOR

UNDERSTANDING LIMITS TO ANURAN DISTRIBUTIONS

Abstract

by Melanie April Murphy, Ph.D.
Washington State University
Aug 2008

Chair: Andrew Storfer

A central objective in ecology and conservation biology is to understand processes limiting species' distributions and population connectivity. This is particularly important for amphibians, which are in global decline at rate exceeding other vertebrates. Niche modeling and landscape genetics are well suited for addressing distribution and connectivity respectively. This dissertation addresses four objectives: i) test whether landscape genetics is possible on spatial and temporal scales relevant to conservation, ii) examine whether spatial processes, environmental condition or dispersal limited niche best explains the observed differences in *Pseudacaris maculata* and *Bufo boreas* distributions in Yellowstone, iii) test alternative hypotheses of ecological processes driving *B. boreas* connectivity in Yellowstone, and iv) develop a new application of gravity models to estimate metapopulation connectivity for *Rana luteiventris* in central Idaho. In *Chapter 1*, I developed and evaluated "genetic surfacing", a continuous method for representing multilocus genetic variation. I detected landscape genetic structure on a contemporary time scale relevant to conservation questions (≥5 generations post

vicariance, migration probability ≤ 0.10), even when population differentiation was minimal ($F_{ST} \geq 0.00015$). Using spatial distribution models in *Chapter 2*, I found environmental conditions limiting species' distributions to be divergent and distance limited niche theory best explained observed species' distributions. In *Chapter 3*, I implemented a novel algorithmic approach to test alternative hypotheses of processes limiting connectivity in *B. boreas*. At fine scales, connectivity was limited by cover, precipitation and roads, whereas ridges, temperature, and precipitation limited connectivity at a broad scale. Using newly derived gravity models in *Chapter 4*, I found *R. luteiventris* connectivity was a function of both at site and between site landscape processes. Primary productivity and fish presence at sites limited production of potential migrants. Temperature and major topographic complexity between sites limited connectivity. The impact of temperature-moisture regimes on all three species suggests that future climate change may have a dramatic impact on anuran distribution and connectivity. The methods developed in this dissertation could be used to predict species' distributions (niche models) and resulting connectivity (landscape genetics) in future landscapes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEDICATION

To my husband,

who provided wonderful scientific conversations,

emotional support, and the opportunity to catch big fish on wild rivers.

May we find “the big one” this summer.

And to my parents,

who thought trudging through wetlands

come snow, sleet, hail, or (sometimes) shine

and the occasional wildlife encounter

was a fun way to spend two summer vacations.

May we eat many more bison together after a day the field.

Thank you

INTRODUCTION

A central objective in ecology and conservation biology is to understand processes limiting species distribution and population connectivity. Ecology often focuses on the question, "Why is a species here?" However, an equally important question is, "Why is this species *not* here?" Species' ranges are generally limited by biotic and abiotic factors, but species are often patchily distributed within their range with varying connectivity among occupied sites (Ovaskainen and Hanski 2004). An explanation is that key abiotic variables (e.g., moisture, elevation) or inaccessibility exclude the species from a particular area. Therefore, on a local or regional scale, landscape complexity is likely critical in limiting dispersal and persistence (Seburn et al. 1997, Coulon et al. 2004, Geffen et al. 2004). This leads to two questions. First, **why are some species more patchily distributed than others?** Second, **in these patchy distributions, what limits connectivity?**

In addressing these questions, detailed consideration of landscape processes is at the core of our understanding of fine scale species' distributions and population connectivity. Processes limiting species distributions can be elucidated using niche theory and species distribution models. Under the available set of environmental conditions (Ackerly 2003), divergent fine-scale distributions could result from random spatial process, differences in suitable environmental conditions ("Grinellian" niche theory), or restricted occupancy of the suitable environment due to limited dispersal ("dispersal limited" niche theory) (Pulliam 2000). Landscape genetics is an emerging discipline that provides a conceptual framework to address connectivity, through an integration of landscape ecology and population genetics (Manel et al. 2003, Holderegger and Wagner 2006, Storfer et al. 2007). By integrating

methods from spatial statistics, hypotheses of connectivity can be tested with multivariate spatial analyses with no *a priori* definitions of relationships among variables.

Understanding limits to species' distribution and connectivity is particularly important for amphibians, as many species are in decline with habitat loss as the main causal factor (Storfer 2003, Stuart et al. 2004). Amphibian niches are likely restricted by environmental conditions such as roads (Mazerolle 2004, Arens et al. 2007, Eigenbrod et al. 2008), fish introduction (Pilliod and Peterson 2001, Bosch et al. 2006), wetland disturbance (Gray et al. 2004), and forest management (Walters 2007). Amphibians also demonstrate population structure on geographic scales appropriate for landscape level analyses of connectivity (Scribner et al. 2001, Palo et al. 2004, Funk et al. 2005, Spear et al. 2006, Arens et al. 2007). In addition, pond-breeding anurans generally have a biphasic lifecycle and therefore utilize both terrestrial and aquatic environments. Finally, identifying landscape variables influencing population connectivity in protected areas is vital for understanding currently unexplained declines (Collins & Storfer 2003).

This dissertation is a study of amphibian distribution and connectivity in two natural areas, Yellowstone National Park and the Bighorn Crags ID. I address four objectives: i) test whether landscape genetics is possible on spatial and temporal scales relevant to conservation (Chapter 1), ii) examine whether spatial processes, environmental condition or dispersal limited niche best explains the observed differences in *Pseudacaris maculata* and *Bufo boreas* distributions in Yellowstone National Park (Chapter 2), iii) test alternative hypotheses of ecological processes driving *B. boreas* connectivity (Chapter 3), and iv) use gravity models to estimate metapopulation connectivity for *Rana luteiventris* in the Bighorn Crags (Chapter 4).

OVERVIEW OF CHAPTERS

Chapter 1: Representing genetic variation as continuous surfaces: An approach for identifying spatial dependency in landscape genetic studies

Chapter 1 addresses three questions: i) Can "genetic surfaces" accurately represent landscape genetic structure?, ii) Is landscape genetics possible on a temporal and spatial scale relevant to conservation? and iii) What is the advantage of a continuous representation of population structure over more global measures of population structure?

Landscape genetics studies are typically focused on explaining the influence of landscape factors on the spatial distribution of genetic variability (Storfer et al. 2007). Current analytical methods often rely on definition of discrete population units, however many landscape processes are spatially continuous (e.g., elevation, moisture) limiting the applicability of delineated populations for landscape genetics studies (Manel et al. 2003, Storfer et al. 2007). In addition, raw data from neutral markers are not direct measurements of a landscape process and must be converted into a spatially informed response variable (e.g., migration rate or a measure of relatedness) (Storfer et al. 2007). To address these issues, I developed and evaluated a method for creating a continuous surface from microsatellite DNA. I then used the derived surface to test the spatial dependence of observed genetic structure in relation to landscape variables.

Chapter 2: Frog and Toad (not) Together: Quantifying differences in species niche and distribution

Chapter 2 addresses three questions: i) are observed distributions of *P. maculata* and *B. boreas* significantly different?, ii) can *P. maculata* and *B. boreas* niche be separated based on

environmental requirements (Grinellian niche)? and iii) are the observed distributions best explained by a Grinellian or dispersal limited niche theory? In answering these questions, I developed modelling and evaluation methodology addressing current limitations of niche-based species distribution models (Guisan and Thuiller 2005, Araujo and Guisan 2006).

The *Bufo boreas* - *Pseudacris maculata* system in Yellowstone is ideal for comparison of niche-based species distribution models (Guisan and Zimmermann 2000, Guisan and Thuiller 2005). *Bufo boreas* and *P. maculata* share several basic habitat requirements including semi-permanent wetlands and warm shallow waters for rearing (Koch and Peterson 1995). However, *P. maculata* is widespread, with a more continuous distribution (Koch and Peterson 1995, Patla et al. 2006) than *B. boreas* which is in decline throughout large portions of its range (Carey 1993, Koch and Peterson 1995, Muths et al. 2003). In addition, *P. maculata* has shorter dispersal capabilities and smaller clutch sizes than *B. boreas* (Spencer 1964, Muths 2003). Differences in environmental requirements and dispersal capabilities may explain the observed difference in distributions.

Chapter 3: Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics

Chapter 3 addresses three questions regarding *B. boreas* connectivity in Yellowstone National Park: i) what ecological processes (habitat permeability, topographic morphology, and temperature-moisture regime) influence connectivity of *B. boreas* breeding sites?, ii) at what scale(s) do these ecological processes affect *B. boreas* connectivity?, and iii) is *B. boreas* connectivity hierarchical with metrics operating at coarser spatial and temporal scales

driving connectivity between genetic clusters; while metrics operating at finer spatial and temporal scales will drive connectivity within a genetic cluster?

Boreal toads (*Bufo boreas*) in Yellowstone National Park are highly suitable for testing ecological processes driving connectivity as measured with neutral genetic data with an ecoinformatic approach (Chon and Park 2006). Distance is a limiting factor for *B. boreas* connectivity (Muths 2003, Manier and Arnold 2006). However, many landscape factors may have an additional affect. *Bufo boreas* is known to be sensitive to habitat permeability as measured by canopy cover (Bartelt et al. 2004), roads (Arens et al. 2007), and thermal influence (Koch and Peterson 1995). Topographically complex areas prevalent throughout Yellowstone may make dispersal physiologically expensive (Lougheed et al. 1999, Funk et al. 2005) and temperature-moisture regimes likely limit connectivity (Palo et al. 2003, Bartelt and Peterson 2005). Each of these processes (habitat permeability, topographic morphology, and temperature-moisture) may influence connectivity at multiple spatial scales. In addition, ecological processes likely operate in a hierarchy with fine-scale processes influencing local connectivity and more broad-scale processes resulting in major breaks in genetic connectivity.

Genetic measures of connectivity may have some advantages over demographically based measures in amphibian studies. Amphibian populations generally fluctuate greatly in census size from year to year, consequently reducing statistical power to detect population trends unless monitoring is long-term (i.e., up to twenty years) (Alford and Richards 1999, Green 2003). In addition, directional migration rates (Pritchard et al. 2000), effective population size (Xu and Fu 2004), and population decline (Cornuet and Luikart 1997, Luikart and Cornuet 1998, Luikart et al. 1998a, Luikart et al. 1998b, Beebee and Rowe 2001) can be

estimated from genetic data in a relatively short time-frame with fewer logistical constraints than demographic studies.

Chapter 4: Frogs in the high mountains: application of gravity models for landscape genetics

Chapter 4 develops gravity models for landscape genetic application and addresses three primary questions: i) is genetic structure clustered by basin or drainage?, ii) is gene flow influenced by both at site and among site landscape processes?, and iii) what landscape processes explain observed gene flow among sites?

Using traditional population genetic approaches and gravity models adapted for landscape genetics, I investigated fine-scale landscape genetic structure of *R. luteiventris* in eight basins in the Bighorn Crags, Idaho. The Bighorn Crags study area is well suited to addressing questions of metapopulation dynamics with genetic data. Previous research on *R. luteiventris* habitat use identified individual movements to reach hibernation, breeding and summer feeding sites (Pilliod *et al.* 2002). *Rana luteiventris* movements through terrestrial habitats are restricted by moisture conditions (Pilliod et al. 2002) and elevation (Funk et al. 2005). In addition, breeding is limited by the presence of fish (Pilliod and Peterson 2001, Pilliod et al. 2002). Gravity models (Voorhees 1956, Anderson 1979, Willig and Bailey 1979, Fotheringham and O'Kelly 1989), adapted from transportation and economic modeling, allow the inclusion of both at site and between site landscape characteristics to estimate their influence on connectivity.

LITERATURE CITED

Ackerly, D. D. 2003. Community assembly, niche conservatism and adaptive evolution in changing environments. International Journal of Plant Sciences 164:S164-S184.

Alford, R. A., and S. J. Richards. 1999. Global amphibian declines: a problem in applied ecology. Annual Review of Ecological Systems 30:133-165.

Anderson, J. E. 1979. A theoretical foundation for the gravity equation. American Economic Review 69:106-116.

Araujo, M. B., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography 33:1677-1688.

Arens, P., T. van der Sluis, W. P. van't Westende, B. Vosman, C. C. Vos, and M. J. M. Smulders. 2007. Genetic population differentiation and connectivity among fragmented Moor frog (*Rana arvalis*) populations in The Netherlands. Landscape Ecology 22:1489-1500.

Bartelt, P., and C. R. Peterson. 2005. Physical modeling operative temperature and evaporative water loss in amphibians. Journal of Thermal Biology 30:93-102.

Bartelt, P., C. R. Peterson, and R. W. Klaver. 2004. Sexual differences in the movements and habitats selected by Western toads in southeastern Idaho. Herpetologica 60:455-467.

Beebee, T., and G. Rowe. 2001. Application of genetic bottleneck testing to the investigation of amphibian declines: a case study with Natterjack toads. Conservation Biology 15:266-270.

Bosch, J., P. A. Rinconi, L. Boyero, and I. Martinez-Solano. 2006. Effects of introduced salmonids on a montane population of Iberian frogs. Conservation Biology 20:180-189.

Carey, C. 1993. Hypothesis concerning the causes of the disappearance of boreal toads from the mountains of Colorado. Conservation Biology 7:355-362.

Chon, T.-S., and V.-S. Park. 2006. Ecological informatics as an advanced interdisciplinary interpretation of ecosystems. Ecological Informatics 1:213-217.

Cornuet, J. M., and G. Luikart. 1997. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001-2014.

Coulon, A., J. F. Cosson, J. M. Angibault, B. Cargnelutti, M. Galan, N. Morellet, E. Petit, S. Aulagnier, and A. J. M. Hewison. 2004. Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. Molecular Ecology 13:2841-2850.

Eigenbrod, F., S. J. Hecnar, and L. Fahrig. 2008. The relative effects of road traffic and forest cover on anuran populations. Biological Conservation 141:35-46.

Fotheringham, A. S., and M. E. O'Kelly. 1989. Spatial interaction models: formulation and applications. Kluwer Academic, Dordrecht.

Funk, W. C., M. S. Blouin, P. S. Corn, B. A. Maxell, D. S. Pilliod, S. Amish, and F. W. Allendorf. 2005. Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. Molecular Ecology 14:483-496.

Geffen, E., M. J. Anderson, and R. K. Wayne. 2004. Climate and habitat barriers to dispersal in the highly mobile grey wolf. Molecular Ecology 13:2481-2490.

Gray, M. J., L. M. Smith, and R. Brenes. 2004. Effects of agricultural cultivation on demographics of Southern high plains amphibians. Conservation Biology 18:1368-1377.

Green, D. M. 2003. The ecology of extinction: population fluctuation and decline in amphibians. Biological Conservation 111:331-343.

Guisan, A., and W. Thuiller. 2005. Predicting species distributions: offering more than simple habitat models. Ecology Letters 8:993-1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modeling 135:147-186.

Holderegger, R., and H. Wagner. 2006. A brief guide to Landscape Genetics. Landscape Ecology 21:793-796.

Koch, E. D., and C. R. Peterson. 1995. Amphibians and Reptiles of Yellowstone and Grand Teton National Parks. University of Utah Press, Salt Lake City.

Lougheed, S. C., C. Gascon, D. A. Jones, J. P. Bogard, and P. T. Boag. 1999. Ridges and rivers: a test of competing hypotheses of Amazonian diversification using a dart-poison frog (*Epipedobates femoralis*). Proceedings of the Royal Society B 266:1829-1835.

Luikart, G., and J. Cornuet. 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology 12:228-237.

Luikart, G., J. M. Cornuet, F. W. Allendorf, and W. B. Sherwin. 1998a. Distortion of allele frequency distributions provide a test for recent population bottlenecks. Journal of Heredity 89:238-247.

Luikart, G., W. B. Sherwin, B. M. Steele, and F. W. Allendorf. 1998b. Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. Molecular Ecology 7:963-974.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189-197.

Manier, M. K., and S. J. Arnold. 2006. Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. Proceedings of the Royal Society B 273:3001-3009.

Mazerolle, M. J. 2004. Amphibian road mortality in response to nightly variations in traffic intensity. Herpetologica 60:45-53.

Muths, E. 2003. Home range and movements of boreal toads in undisturbed habitat. Copeia 1:160-165.

Muths, E., P. S. Corn, A. P. Pessier, and D. E. Green. 2003. Evidence for disease-related amphibian decline in Colorado. Biological Conservation 110:357-365.

Ovaskainen, O., and I. Hanski. 2004. Metapopulation dynamics in highly fragmented landscapes. Pages 73-104 *in* I. Hanski and O. E. Gaggiotti, editors. Ecology, genetics, and evolution of metapopulations. Elsevier, Burlington MA.

Palo, J. U., R. B. O'Hara, A. T. Laugen, A. Laurila, C. R. Primmer, and J. Merila. 2003. Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. Molecular Ecology 12:1963–1978.

Palo, J. U., D. S. Schmeller, A. Laurila, C. R. Primmer, S. L. Kuzmin, and J. Merila. 2004. High degree of population subdivision in a widespread amphibian. Molecular Ecology 13:2631-2644.

Patla, D., C. R. Peterson, and R. Bennetts. 2006. Amphibian monitoring in the Greater Yellowstone Network.

Pilliod, D. S., and C. R. Peterson. 2001. Local and landscape effects of introduced trout on amphibians in historically fishless watersheds. Ecosystems 4:322-333.

Pilliod, D. S., C. R. Peterson, and P. I. Ritson. 2002. Seasonal migration of Columbia spotted frogs (*Rana luteiventris*) among complementary resources in a high mountain basin. Canadian Journal of Zoology 80:1849–1862.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pulliam, H. R. 2000. On the relationships between niche and distribution. Ecology Letters 3:349-361.

Scribner, K., J. W. Arntzen, N. Cruddace, R. S. Oldham, and T. Burke. 2001. Environmental correlates of toad abundance and population genetic diversity. Biological Conservation 98:201-210.

Seburn, C. N. L., D. C. Seburn, and C. A. Paszkowski. 1997. Northern leopard frog (*Rana pipiens*) dispersal in relation to habitat. Pages 64-72 *in* D. M. Green, editor. Amphibians in decline: Canadian studies of a global problem. Society of Amphibians and Reptiles.

Spear, S., C. R. Peterson, M. Matacq, and A. Storfer. 2006. Molecular evidence for historical and recent population size reductions of tiger salamanders (*Ambystoma tigrinum*) in Yellowstone National Park. Conservation Genetics 7:605-611.

Spencer, A. W. 1964. The relationship of dispersal and migration to gene flow in the boreal chorus frog. PhD. Colorado State University.

Storfer, A. 2003. Amphibian declines: future directions. Diversity and Distributions 9.

Storfer, A., M. A. Murphy, J. S. Evans, C. Goldberg, S. Spear, S. Robinson, R. Dezzani, E. Demmelle, L. Vierling, and L. P. Waits. 2007. Putting the 'landscape' in landscape genetics. Heredity 98:128-142.

Stuart, S. N., J. S. Chanson, N. A. Cox, B. E. Young, A. S. L. Rodrigues, D. L. Fischman, and R. W. Waller. 2004. Status and trends of amphibian declines and extinctions worldwide. Science 306:1783-1786.

Voorhees, A. M. 1956. A genetial theory of traffic movement. *in* Institute of traffic engineering, New Haven, Connecticut.

Walters, S. 2007. Modeling scale-dependent landscape pattern, dispersal, and connectivity from the perspective of the organism. Landscape Ecology 22:867-881.

Willig, R. D., and E. E. Bailey. 1979. The economic gravity model. American Economic Review 69:96-101.

Xu, H., and Y.-X. Fu. 2004. Estimating effective population size or mutation rate with microsatellites. Genetics 166:555-563.

CHAPTER FORMATS AND ATTRIBUTION

My dissertation is a collection of four manuscripts for publication in scientific journals. Although I made some minor formatting adjustments for the sake of consistency among chapters, each chapter is formatted for its respective journal including reference format and limits on display items (tables and figures). I was the primary contributor for each chapter conducting data collection, analytical development, analysis, and writing. However, both the challenge and strength of landscape genetics is the integrated nature of the field. Therefore, several collaborations were important for integrating cutting-edge ideas and approaches for this dissertation. As my committee chair, Andrew Storfer contributed to idea refinement and manuscript revisions on all chapters and is thus a coauthor on all chapters. I collaborated with Jeffrey Evans, a landscape ecologist, to integrate cutting edge tools from landscape ecology and apply these tools to population genetic data. I collaborated with Ray Dezzani to develop an implementation of gravity models for landscape genetics. Chuck Peterson contributed to the development of sample design and interpretation of results based on years of experience working in Yellowstone. David Pilliod developed the field design, was responsible for sample collection, and provided context for the results based on more than 10 years experience in the Bighorn Crags for Chapter 4. At the time of defense, Chapter 1 was resubmitted and in review for Ecography. Chapter 2 is formatted for Ecology which has a strict page limit and instructs authors to rely heavily on appendices as a way to include additional information. Chapter 3 is in review with Ecology. Chapter 4 is formatted for The American Naturalist, which prefers tables and figures imbedded in the text.

CHAPTER 1 - Representing genetic variation as continuous surfaces: An approach for identifying spatial dependency in landscape genetic studies

MELANIE A. MURPHY[1], JEFFREY S. EVANS[2], SAMUEL A. CUSHMAN[3], ANDREW STORFER[1]

M. A. Murphy (mamurphy@wsu.edu) - School of Biological Sciences, Department of Biology, Washington State University, Pullman WA, 99164, USA. J. S. Evans - USDA Forest Service, Rocky Mountain Research Station, 1221 S. Main, Moscow, ID 83483, USA. S. Cushman - USDA Forest Service, Rocky Mountain Research Station, Missoula, MT, 59801, USA. A. Storfer - School of Biological Sciences, Department of Biology, Washington State University, Pullman WA, 99164, USA.

Resubmitted to Ecography and in review

ABSTRACT

Landscape genetics, an emerging field integrating landscape ecology and population genetics, has great potential to influence our understanding of habitat connectivity and distribution of organisms. Whereas typical population genetics studies summarize gene flow as pair-wise measures between sampling localities, landscape characteristics that influence population genetic connectivity are often continuously distributed in space. Thus, there are currently gaps in both the ability to analyze genotypic data in a continuous spatial context and our knowledge of expected of landscape genetic structure under varying conditions. We present a framework for generating continuous "genetic surfaces", evaluate their statistical properties, and quantify statistical behavior of landscape genetic structure in a simple landscape. We simulated microsatellite genotypes under varying parameters (time since vicariance, migration, effective population size) and used ancestry (q) values from STRUCTURE to interpolate a genetic surface. Using a spatially adjusted Pearson's correlation coefficient to test the significance of landscape variable(s) on genetic structure we were able to detect landscape genetic structure on a contemporary time scale (≥5 generations post vicariance, migration probability ≤ 0.10) even when population differentiation was minimal ($F_{ST} \geq 0.00015$). We show that genetic variation can be significantly correlated with geographic distance even when genetic structure is due to landscape variable(s), demonstrating the importance of testing landscape influence on genetic structure. Finally, we apply genetic surfacing to analyze an empirical dataset of black bears from Northern Idaho, USA. We find black bear genetic variation is a function of distance (autocorrelation) and habitat patch (spatial dependency), consistent with previous results indicating genetic variation was influenced by landscape by resistance. These results suggest

genetic surfaces can be used to test competing hypotheses of the influence of landscape characteristics on genetic structure without delineation of categorical groups.

INTRODUCTION

A key objective of ecological studies is to understand the influence of biotic and abiotic factors on population connectivity, and resulting fine-scale species' distributions across a landscape. As a relatively new approach to assess connectivity, landscape genetics is an emerging discipline that aims to quantify the effect of landscape composition, configuration and matrix quality on the spatial distribution of genetic variation (Holderegger and Wagner 2006, Storfer et al. 2007). Evaluating neutral genetic variation in a landscape context has already provided insights into species' ecology such as identification of potential barriers to gene flow (for review, see Manel et al. 2003, Storfer et al 2007). However, our ability to quantify the relationship between genetic variation and multiple landscape variables with a robust assessment of error is currently constrained by the unique nature of multilocus data and available analytical tools (Storfer et al. 2007).

Compared to typical ecological data, spatial incorporation of neutral multilocus genetic data presents two unique complexities. Whereas typical ecological response variables (e.g., soil moisture, tree height, site occupancy) have direct ecological interpretation and the observed value can be input directly into a statistical model, neutral genotypic measurements such as microsatellites are a collection of DNA fragment lengths. These data are only a meaningful response variable when being considered in the context of relationships among alleles, as allele frequency distributions, and/or differences in heterozygosity. In addition, ecological measurements can be associated with explicit spatial locations. In contrast, genetic summary

statistics are generally not associated with sample locations, but represent an aspatial genetic distance between sample pairs (e.g, F_{ST}, Nei's distance).

Several studies propose methods to estimate spatial location(s) of gene flow between pairs of sample locations (Arnaud 2003, Cushman et al. 2006, McRae 2006, McRae and Beier 2007, Spear et al. 2005). These approaches are extremely valuable; however, most methods only select a single path through the landscape based on ranked costs (but see McRae and Beier 2007). Although multiple costs or combinations can be tested (Cushman *et al.* 2006), empirical costs of movement though the landscape may be unknown, difficult to hypothesize, or landscape variables may have a non-linear relationship with gene flow. In addition, significance of landscape genetic variation is often evaluated using the partial Mantel tests where significance tests may be unreliable (Castellano and Balletto 2002, Raufaste and Rousset 2001, Rousset 2002).

As an alternative, spatial autocorrelation statistics have been used to explain genetic variation in a spatially continuous manner (Epperson 2003, Shimatani and Takahashi 2003, Slatkin and Arter 1991, Sokal et al. 1997, Sokal et al. 1998). However as opposed to strict spatial autocorrelation, the pattern of genetic variation may be the result of spatial dependency (Wagner and Fortin 2005), where another spatial variable drives genetic variation (e.g., habitat patch; Bockelmann et al. 2003, Keyghobadi et al. 2005). If this variable is autocorrelated, the result can be a significant autocorrelation statistic, deemed "false autocorrelation" (Legendre et al. 2002). In addition, there may be an interaction between autocorrelation and spatial dependency (Fortin and Dale 2005, Legendre et al. 2002). An optimal solution would allow testing of multiple landscape characteristics with competing hypotheses of autocorrelation, spatial dependency, and autocorrelation-spatial dependency interaction without the necessity of defining costs of

independent variables (Storfer et al. 2007). To achieve this goal, raw multilocus genotypic data could be converted into point representations of genetic variation, creating a common theoretical base for data analysis between population genetics and spatial statistics (Shimatani and Takahashi 2003).

Representing whole genotypes as spatially referenced, continuous measures of genetic structure would advance our ability to test the effect of continuous landscape variables with well-developed multivariate spatial statistics, such as spatial regression methods (Fotheringham et al. 2002, Haining 2003) and point pattern analysis (Diggle 2003). These methods offer several advantages relative to widely applied methods in landscape genetics including: ability to estimate the influence of multiple independent variables simultaneously, parameter estimates for these variables that are valid in the presence of spatial autocorrelation, robust assessment of uncertainty, and ability to use landscape genetic models for spatial prediction (Wagner and Fortin 2005). Thus far, such methods have yet to be derived for landscape genetics.

To address these issues, we develop and evaluate an approach based on a novel integration of available methodological components for creating a continuous surface of genetic variation. We use a Bayesian clustering algorithm (Pritchard et al. 2000) to generate an ancestry value for each spatially referenced genotype to generate point values for surface interpolation. Via simulation, we address the following questions using a genetic surfacing approach: 1) Does landscape genetics have the power to detect the effect of contemporary landscape condition?; 2) How much data are needed to have the power to detect landscape genetic structure?; 3) Can landscape genetics accurately identify the process(es) generating landscape genetic structure?; and 4) What is gained using a landscape genetic approach compared to standard global statistics (e.g., F_{ST})?

Finally, we use a black bear genetic dataset from Northern Idaho, USA (Cushman et al. 2006) to give an example of genetic surfacing with an empirical dataset.

METHODS

Simulation Components

Our genetic surfacing simulations had six main steps: defining the landscape, simulation of multilocus genotypes, subsampling genotypes, estimating ancestry, surface interpolation, and testing models of landscape genetic structure. We outline the steps of the simulations below and as a flow chart (Fig. 1). Additional details are located in Appendix 1.

Step 1 – Defining the Landscape

We defined two landscapes (48 km X 48 km) in ArcInfo (ESRI 2005). Landscape A (Fig. 1A) contained two patches of equally suitable habitat separated by unsuitable habitat and was used to simulate a landscape effect independent of distance (i.e., spatial dependency). In landscape B (Fig. 1B), each patch was partitioned into five subpatches of equally suitable habitat for simulations including autocorrelation. In an empirical landscape, habitat patches could be defined by landscape variables (e.g., slope or moisture) or ideally, the landscape could be represented by continuous values (McGarigal and Cushman 2005).

Step 2 – Simulation of Multilocus Genotypes

To address if landscape genetics has the power to detect the effect of current landscape condition, we simulated microsatellite genotypes in EASYPOP 1.8 (Balloux 2001) under various conditions (Table 1). In EASYPOP we parameterized either two (Fig. 1A; one "population" per habitat patch) or 10 "populations" (Fig. 1B; one "population" per subpatch). It should be noted that these are "populations" only as the *population* parameter in EASYPOP (Balloux 2001). That

is, with high levels of migration, they may be genetically indistinct. In landscape A, we varied time since vicariance (T, number of generations of separation), migration (M, probability per individual of migration per generation), and effective population size (N, number of breeding individuals contributing to the next generation) (Balloux 2001). In landscape B, we implemented a hierarchical stepping-stone model of migration (HM) (Balloux 2001). The hierarchical stepping-stone model of migration has two migration parameters: migration rate between subpatches within the same patch and migration rate between patches (Fig. 1B). We used this model to simulate a distance effect (i.e., autocorrelation without an effect of the unsuitable habitat) and an interaction between distance and landscape patch (Fig. 1B, see Table 1 for range of conditions). To model genetic variation within populations of a mobile species, we incorporated the potential for individuals to move freely within their given habitat patch within a generation. We implemented this by randomly assigning individuals within their patch (landscape A, Fig. 1) or subpatch (landscape B, Fig. 1) using ArcMap 9.1 (ESRI 2005) for each simulation.

Step 3 – Subsampling

To address how much data are needed to detect landscape genetic structure, we varied sample size and number of loci used by subsampling genotypes and loci of each simulated dataset. Using R statistical package (R Development Core Team 2006), we randomly sampled genotypes without replacement (at 100, 20, 10, and 5%) across the entire landscape to emulate different levels of field sampling without knowledge of habitat patch or population boundaries. We then analyzed each subsample for four quantities of loci: all 30, 20, 10 and 5 loci.

Step 4 – Estimating Proportion of Ancestry Values

We used the Bayesian clustering algorithm STRUCTURE (Pritchard et al. 2000) to derive a continuous measure of genetic variation based on unclassified ancestry values (ranging from 0 - 1). STRUCTURE ancestry values are a proportion of ancestry of each individual's genotype to each of *K* populations ("genetic clusters"). A gradient in genetic structure is represented by expressing the degree of ancestry in a given genetic cluster, maintaining the spatial and statistical variability within cluster. This is akin to fuzzy set theory as applied in remote sensing applications (Bosserman and Ragade 1982, Metternicht 2003), where the response variable may be proportion or probability of membership in multiple habitat classes. The number of clusters identified by STRUCTURE was selected using ΔK (Evanno et al. 2005, Pritchard et al. 2000), and appending ancestry values to each genotype in R (Fig. 1).

Step 5 – Interpolation of a Genetic Surface

We interpolated a genetic surface from the ancestry values for each observed genotype using analyses that are comparable, repeatable, and automated across conditions and replicates. In no cases did STRUCTURE identify more than two clusters (Evanno et al. 2005); therefore, one genetic surface could be generated from the ancestry values for each individual for one of the two genetic clusters. In addition, to quantify whether random effects can be correctly identified with a genetic surface (i.e., no structure), surfaces were interpolated from ancestry values for all data sets (K=2). We created genetic surfaces with an inverse distance weighted interpolation (IDW) (Cressie 1993) with the following parameters: a power function of 2 to control the tension of the surface, a variable search radius of 12 neighboring observations, and a resolution of 30 m^2.

Step 6 – Test Landscape Genetic Models

Under our simulated conditions, genetic variation could be due to spatial dependency ("habitat patch"), spatial autocorrelation ("distance"), both ("patch-distance interaction") (Legendre et al. 2002), or random effects. We calculated a raster to represent each hypotheses of landscape genetic structure in ArcInfo (ESRI 2002). The habitat patch raster consisted of a unique value assigned to each habitat patch (Fig. 1). The distance raster was mean distance (m) for a subpatch from patch 1a as a reference point (Fig. 1). The patch-distance interaction raster evaluated an interaction between habitat and distance. We achieved this by multiplying distance from the unsuitable habitat by patch value (1 for patch 1 and -1 for patch 2, Fig. 1). Using a modified version of program MODTTEST (Legendre 2000), we calculated a Pearson's correlation coefficient between the genetic surface and each of the three models of landscape genetic structure under each simulation condition (n=1000) (Dutilleul 1993, Fortin and Payette 2002, Legendre et al. 2002). As implemented in MODTTEST, in our case the Pearson's correlation is insensitive to violations of bivariate normal in the landscape patch analysis (r=0.987 to Spearman-rank correlation).

Power, Accuracy and Process Identification

We assessed model performance by power, accuracy, and identification of the simulated process (distance, habitat patch, patch-distance interaction). We assessed statistical power (1-β) for both the complete simulated dataset and subsampling levels from Step 3. This addresses both the power to detect landscape genetic structure for a given set of conditions, as well as the amount of data (sample size and number of loci) required to detect that landscape genetic structure.

We evaluated accuracy by two methods: presence of Type I errors and correct identification of simulated landscape genetic structure (distance, habitat patch, or patch-distance interaction). Type I errors occur where significant genetic structure is detected in the absence of simulated genetic structure (Legendre et al. 2002, Zar 1999). We assessed Type I error under two conditions: 1) genotypes with randomized XY coordinates and 2) simulation conditions where no landscape genetic structure was present. To assess the first, we randomized genotypes in step 2 and then completed the analysis twenty times to evaluate a condition known to have no significant landscape genetic structure (Fig. 1). For the second assessment of Type I error, we identified simulations where the null hypothesis (no landscape genetic structure) was true (T1; simulations where no significant landscape genetic structure was detected using the complete dataset, see Table 1). If landscape genetic structure is not detected with the complete dataset, any subsample with a significant result is a Type I error.

To evaluate the accuracy of variable selection, we calculated the number of instances where the simulated ("correct") model of landscape genetic structure was identified out of the total number of analyses. As a general estimate of model choice, we calculated the number of instances where the "correct" relationship had the highest correlation coefficient out of the total number of significant analyses. For simulations in landscape A, the "correct" relationship was the habitat patch model (Table 1). For simulations in landscape B, the "correct" relationship was the distance model for HM4 and HM8 where migration was equal across the landscape (Table 1).

For the remaining simulations, identifying the "correct" relationship was more complex because we simulated multiple relationships simultaneously. We identified the "best" relationship based on the ability to detect landscape genetic structure in simulations from landscape A (Appendix 1). For example, if the within patch migration probability is 0.2 and

between patch migration 0.1, we would not expect to detect a distance effect if 0.2 probability of migration was undetectable in landscape A. Therefore, the "best" relationship in these simulations would be the habitat patch model. In addition, when the migration probability was disparate between versus within habitat patches, genetic surfacing may only detect the stronger restriction to migration probability. We considered the strongest correlation non-spurious as long as it identified either the "best" relationship or the landscape genetic model corresponding to the lowest migration probability. Then we evaluated the power to detect a significant interaction between autocorrelation and spatial dependency (i.e., patch-distance interaction) for different levels of migration within and between patches.

Comparison to Standard Genetic Statistics

To evaluate the comparability of the genetic surfacing approach to standard genetic statistics, we calculated F_{ST} (standard measure of genetic distance based on heterozygosity) and allelic richness by patch for all simulations. We then calculated the number of significant genetic surfaces by level of genetic differentiation (F_{ST}) across all simulations.

Empirical Illustration

We utilized a previously published empirical dataset of black bear (*Ursus americanus*) genotypes from northern Idaho, USA (Cushman et al. 2006) for a simple illustration of the genetic surfacing approach. The dataset consists of 146 unique black bear genotypes generated using nine microsatellite loci collected on a 2.6 km^2 grid (Cushman et al. 2006). We applied the genetic surfacing framework, with minor adjustments. So that results would be comparable with out simulations, we identified relatively coarse-scale habitat patches based on the major elevation break in a digital elevation model (DEM) for the study area (elevation = 800 m) determined by presence of suitable habitat and bear observations. This created four habitat

patches (Fig. 2): suitable habitat (1, 2, 4) and unsuitable habitat (3). Patch four (Fig. 2) was outside the study area; therefore it was excluded from the analysis. We used the resulting patch configuration to create the habitat patch, distance and patch-distance interaction models as in the simulation (see Fig 1, Step 6). We executed STRUCTURE, interpolated the genetic surface, and performed the spatially adjusted Pearson's correlation as in the analyses of the simulated data (Fig. 1). We then compared the results to those of Cushman et al. (2006).

RESULTS

Using genetic surfacing, we were able to detect landscape genetic structure generated by contemporary landscape condition (≥5 generations post vicariance, Table 1), with low to moderately high levels of migration (migration probability 0.0-0.1) and when population differentiation is minimal ($F_{ST} \geq 0.00015$). In addition, we demonstrate that the correlation between genetic variation and distance can be significant when the simulated process was habitat patch only (i.e., spatial dependency). However, by comparing multiple models (habitat patch, distance, patch-distance interaction) we were able to correctly identify the simulated relationship (e.g., habitat patch alone) 96% of the time (Appendix 2).

We found genetic surfacing to be insensitive to misidentification of number of genetic clusters (*K*) in STRUCTURE. In all simulations, analyses in STRUCTURE suggested that either one or two genetic clusters (*K*) were most likely. However, in 9.37% (322/3434) of analyses, *K*=2 was not a significant improvement over *K*=1 (Table 1, Appendix 2). Even still, we found evidence of continuous genetic structure. We identified significant landscape genetic structure for 16.15% (52/322) of these analyses with moderately high power (0.7453), and low Type I error (3.73%, 12/322). In addition, when ancestry values for *K* greater than two were use for

surface interpolation for a cross section of simulations, Pearson's correlations and model choice did not significantly change (r = 0.91).

Statistical Power – Time, Migration, and Effective Population Size

Using genetic surfacing, we were able to detect the effect of current landscape condition five generations post vicariance, 0 - 0.1 probability of migration, for the range of simulated population sizes, and in the presence of internal patch structure (Table 1, Appendix 1). The strength of the relationship between the landscape model and the genetic surface increased as time since vicariance increased from zero, significant for all simulations 5 – 1000 generations post vicariance (Table 1; Figs. 3a,b). Genetic surfaces appeared to be less distinct in the presence of migration between habitat patches (Figs. 3c,d). However, all analyses detected significant landscape genetic structure for simulations with migration probabilities between 0 (M1) – 0.1 (M6) (r = 0.574; Table 1). Variation in effective population size or unequal population size had little effect on the power to detect the simulated genetic structure, unless global effective population size was small (n=50) compounded by the presence of migration (Appendices 1,2). In the presence of internal patch structure (Fig. 1B), we were able to detect the simulated landscape genetic structure 97.5% of the time (Table 1).

Statistical Power – Data requirements

We gained more power by increasing sample size as opposed to increasing number of loci (Fig. 4a), with data requirements dependent on the amount of genetic structure. As we increased time since vicariance, fewer samples were necessary to have sufficient power to detect significant landscape genetic structure (Fig. 4b). For example, with 20% sampling genetic surfacing had 0.90 power 50 generations post vicariance but only 0.20 power to detect landscape genetic structure 10 generations post vicariance (Fig. 4c). At the same sampling level, when

migration probability increased from 0.05 to 0.1, power decreased from 0.92 to 0.10 (Appendix 2). However, power increased to 0.55 with complete sampling of simulations with migration probability of 0.1 (Fig. 4d). For more moderate levels of migration (≤0.005), we were able to detect landscape genetic structure regardless of sample size or number of loci (Fig. 4d, Appendix 2). In general, population size and unequal population size had little effect on power (for details, see Appendices 1,2).

Accuracy and Process Identification

We were able to identify the presence of landscape genetic structure and the model used to simulate landscape genetic structure accurately. In the randomization assessment of accuracy, only 3.8% (468/12,300) of the correlations between the genetic surface and model of landscape genetic structure were significant, below the selected p-value (0.05). The Type I error rate was also below 5% for all samples sizes and number of loci (0.11 – 0.48). In simulation conditions without landscape genetic structure (T1, T2, M7), the Type I error was 5% (12/240). When evaluating the accuracy to identify the simulated process (habitat patch, distance, patch-distance interaction), all correlations between the genetic surface and each model of landscape genetic structure was often significant (Figs. 5 a,b,c). After model comparison the "correct" model was selected for most simulations without internal patch structure (>95%; Appendix 2, Figs. 5d,e). When we simulated a distinct hierarchical effect (HM5), genetic surfacing selected the patch-distance interaction model ("best" model; Fig. 5f; Appendix 2). When the hierarchical effect was less marked (HM1, HM6) or within patch migration was high (HM2, HM3, HM6, HM7) the habitat patch model was selected (Figs. 5c,f; Appendix 2). When migration probability was constant between subpatches across the entire landscape (HM4, HM8), the distance model was correctly chosen more frequently than the other two models (Appendix 2).

Comparison to Standard Genetic Statistics

Across all simulated conditions and sample sizes, we were able to detect landscape genetic structure even when F_{ST} values were minimal (0.00015 - 0.05, r = 0.113 – 0.934, p-value <0.001; Fig. 6). This result was most striking when we simulated internal patch structure, which produced more continuous genetic structure (Appendix 2). Significant landscape genetic structure when F_{ST} was minimal was not the result of Type I errors, which was < 0.05 for these cases.

Empirical Application

We identified weak genetic structure in the black bear samples from the study area with two supported genetic clusters (K =2) and F_{ST} = 0.061 (Pritchard et al. 2000). Raw ancestry values formed a continuous distribution from 0.0806 – 0.909 (mean = 0.482, S.D. = 0.227), analogous to a transition from one cover type to a second cover type (Fig. 2). We had the power to detect a landscape genetic structure, with significant correlations with all three models of landscape genetic structure and the genetic surface (p-values 0.01 - <0.0001). Out of the tested models, the patch-distance interaction model (autocorrelation with spatial dependency) had the most support (r = 0.645; p-value < 0.0001). In contrast, if we had applied STRUCTURE for discrete classification of population membership with an assignment threshold of 0.75 (ancestry value), 63.7% (93/146) of individuals would be unclassified ("admixed") indicating lack of genetic structure.

DISCUSSION

Our approach demonstrates that it is possible to achieve one of the central goals of landscape genetics: detect the effects of the current landscape on genetic variation of a focal species, even

with minimal genetic structure characteristic of fine-scale studies (Coulon et al. 2004). In addition, a genetic surfacing approach addresses some of the current limitations in landscape genetics by providing a statistically powerful, continuous representation of genetic variation. When applied to a published, empirical dataset our approach produced consistent results. Finally, although we used discrete landscape patches for the sake of reproducibility and comparability among simulations, genetic surfacing provides a framework for modeling the effects of continuous landscape variables in empirical application.

Statistical Power – Data requirements

Genetic surfacing had the power to detect landscape genetic structure under conditions likely observed in extant landscapes: ≥ 5 generations post vicariance and zero to moderately high migration rates (Riley et al. 2006, Scribner et al. 2001). This is a substantial improvement over more traditional methods where up to 60 generations have been required to detect landscape effects (Holzhauer et al. 2006, Keyghobadi et al. 2005). Genetic metrics based on heterozygosity (such as F_{ST}) are less sensitive to changes in genetic variation than metrics based on allele frequency distributions (e.g., ancestry values) and could represent the effects of past process in the presence of more recent landscape change (Holzhauer et al. 2006, Keyghobadi et al. 2005).

The data required for high statistical power of genetic surfacing are practical for broad application to empirical systems, given the following considerations. First, our simulations likely underestimate microsatellite allelic diversity relative to empirical application. We randomly subsampled microsatellite loci for analysis, some of which were invariant. However, researchers generally exclude monomorphic loci and may select highly polymorphic loci. Second, adding samples to the analyses increases statistical power of genetic surfacing more quickly than adding loci (Fig. 4a). In application, it is generally easier for researchers to increase

sample size than develop additional loci, which add power in conventional population genetic application. Because we randomly sampled genotypes across the landscape, increasing sample size enhances the spatial distribution of observations improving the estimate (Legendre et al. 2002). Third, the absolute number of samples per population does not notably increase (20 - 30 samples) as population size decreases or populations become unequal in size (Appendix 2). Finally, more complex spatial sampling designs than the random sample employed in this study may reduce the number of samples required for equivalent power (Rempel and Kushneriuk 2003).

Accuracy and Process Identification

Genetic surfacing as presented in this paper is accurate and robust. The low Type I error rates demonstrate that significant relationships between the genetic surface and the three landscape genetic models are not artifacts of the simulation conditions, use of ancestry values, or how landscape rasters were calculated. In addition, we show that landscape genetic structure can be distinguished from random effects and is robust to a misidentification of *K*. However, genetic surfacing is not dependent on STRUCTURE ancestry used in this demonstration. Any genotypic point data calculated independently from space could be used for a surface interpolation including other clustering algorithms (Corander et al. 2003), proportion of shared alleles, or other point estimates of genetic diversity.

Although the genetic surfacing approach is accurate, misidentification of the process generating observed genetic structure is possible when multiple models are not considered. We demonstrate that distance can be significant due to 'false' autocorrelation (Legendre 1993). That is, genetic surfacing detected a significant distance relationship when observations were not spatially autocorrelated but rather dependent on some other variable, which is itself

autocorrelated. For example, although genetic structure was simulated solely as a product of habitat patch in landscape A (Fig.1A), significant correlation coefficients between the genetic surface and the distance model are present due to autocorrelation of habitat patches ($I = 0.98$, $p<0.001$). However, when we compared multiple models, habitat patch explained more variation than distance or patch-distance interaction. This demonstrates that genetic surfacing can differentiate between spatial autocorrelation (i.e., isolation-by-distance) and spatial dependency (i.e., genetic structure that is dependent on an autocorrelated variable such as habitat patch) (Fortin and Dale 2005). This result emphasizes the value of incorporating landscape variables in broader population genetics research and the importance of testing multiple hypotheses of genetic structure.

Genetic surfacing did not always detect patch-distance interaction as the correct condition for data generated under this model (Appendix 2). Distance was only identified if within patch migration was restricted (probability of migration = 0.05) but not extremely different from the migration probability between habitat patches (Fig. 5c). There are two principal reasons for this limitation. First, genetic divergence with high levels of migration within patches (especially with low levels of migration between patches) is typically low. Second, in the stepping-stone simulations with internal patch structure, the Pearson's correlation may have low power to distinguish between the three landscape models (distance, habitat patch, patch-distance interaction). Since genetic surfacing identified hierarchical structure when the simulated hierarchical genetic structure was most distinct (HM5), it is likely genetic signal that limits model choice more than more than spatial signal. Significance testing via a randomization method (Gardner et al. 1987) may improve model choice.

Comparison to Standard Genetic Statistics

Recent work suggests low values of F_{ST} may be statistically, but not necessarily biologically, significant when using highly polymorphic genetic markers such as microsatellites (Hedrick 2005, Waples and Gaggiotti 2006). With high power attained with highly polymorphic loci, it is possible to detect significant subdivision (i.e., Fst is significantly > 0) when the actual value of the statistic is quite low (e.g.,. $F_{ST} < 0.05$) and the implied level of gene flow is quite high. However, our continuous genetic surfacing approach suggests ecologically relevant landscape effects may be present even when $F_{ST} <0.05$ (Fig. 6). Because the genetic surface is a spatially continuous measure of genetic structure, it reflects population structure not captured in a global summary such as F_{ST}. Thus, analyzing the genotypic data in a spatially informed manner and using spatial analysis methods increases the power to explain how landscape features influence population genetic structure relative to standard methods.

Empirical Application

In application to an empirical dataset, we demonstrate the ability to detect landscape genetic structure with a realistic amount of data (146 samples, 9 loci) and select an ecologically reasonable model using a dataset with minimal global genetic structure. In analysis of this dataset, Cushman et al. (2006) tested multiple hypotheses of genetic differentiation, including isolation-by-distance, barrier (valley), and landscape resistance. They found the landscape resistance model had the most statistical support of their tested models. We tested three coarser –scale models of landscape genetic structure consistent with our simulations: distance (analogous to the isolation-by-distance model), habitat patch (analogous to the barrier model), and patch-distance interaction. We found a habitat patch-distance interaction to have the most support. Of our tested models, this is the only model that incorporates distance and habitat influences, both

of which Cushman et al (2006) incorporate in the landscape resistance model. Low elevation, water, and non-forest are present in the "unsuitable habitat" of our patch model and have high associated resistance in the landscape resistance model. In addition, the landscape resistance model incorporates a measure of autocorrelation by calculating cumulative costs between pairs of observations.

Our analysis of the black bear dataset was a substantial improvement over a standard population classification approach. If we used a conventional application of STRUCTURE and classified individuals into discrete "populations", the number of unclassified individuals (63.7%) would lead to the conclusion of little structure in these data. By treating genetic structure as a continuous surface, and thus the ability to represent gradients of genetic structure, we found an increased sensitivity to spatially dependent relationships compared to a discrete classification of black bears. Finally, although we used a continuous representation of genetic variation, we classified landscape variation into discrete patches to be consistent with the simulated conditions. In future genetic surfacing applications, continuous representation of both landscape and genetic variation would be more powerful allowing for both genetic and environmental gradients (McGarigal and Cushman 2005, Shimatani and Kubota 2004).

Future Directions and Conclusions

Genetic surfacing provides a common statistical framework for population genetics, spatial ecology, and spatial statistics for data analysis. In future applications, the wide range of available multivariate spatial statistical techniques can be applied with genetic surfacing (Haining 2003). These multivariate models will allow researchers to quantify the effect of multiple landscape variables simultaneously for a more complete understanding of how ecological variables affect species' distributions. In addition, the topology of the genetic surface

itself may provide meaningful information. For example, genetic surfaces can be analyzed in a similar manner to other surfaces to quantify topographic complexity (Mucina et al. 1991), spatial trend (Lichstein et al. 2002), or identify areas of discontinuity and their associated landscape variables.

Application of more sophisticated surfacing methods than IDW, implemented for reproducibility and comparability amount simulations, may increase the power of genetic surfacing. This is especially true in the presence of complex genetic structure resulting from multiple interacting landscape processes. Alternative methods that can produce a surface incorporate localized variation in the data such as empirical semivariogram/kriging models (Cressie 1986, Yfantis et al. 1987) or can apply more complex curve fitting parameters in spline models (Mitasova and Jaroslav 1993). In addition, alternative methods allow for the inclusion of an estimate of uncertainty in the original point locations**.** Given multiple locations or error in geographic positioning system (GPS) locations, kernel density estimates can give a probability density function of location over a range (Wand and Jones 1995). However, it is important to understand that surfacing methods estimate a response variable across the extent and assume a well-distributed spatial sample (Lam 1983, Tobler 1979) and point data could be used directly without surface interpolation (Boots and Getis 1988, Diggle 2003). In addition, due to the complexity of data collected in natural systems, low to moderate correlations between genetic structure and spatial variables are possible by chance alone. Simulation or neutral landscape tests may be necessary to establish the significance of results (Isaaks and Srivastava 1989, Lancaster 2006).

In all of our analyses, we were able to use K= 2 and therefore construct a single genetic surface representing membership in both genetic clusters. Although STRUCTURE is not the only

potential implementation of genetic surfacing, in future application based on STRUCTURE there will be cases where more than two genetic clusters are supported. Several approaches may be employed in cases where $K > 2$. A genetic surface, and model of landscape genetic structure, could be constructed for each of K-1 clusters. This may be extremely useful in cases where researchers suspect different processes explain genetic variation for different genetic clusters across the landscape. Alternatively, genetic surfacing could be applied with methods that allow for multiple response variables. Fuzzy set theory allows for the response variable to have proportional membership in multiple classes, which could be modeled as a single response variable (Bosserman and Ragade 1982, Metternicht 2003).

Genetic surfacing will help fill a gap between population genetics and spatial ecology. The approach integrates landscape and genetic analyses to allow testing of multiple ecologically relevant hypotheses (autocorrelation, spatial dependency, or an interaction) without the necessity of analyzing pairwise distance data. In addition, this study demonstrates that significant distance relationships may be the result of spatial dependency or 'false' autocorrelation (Legendre et al. 2002), emphasizing the importance of considering landscape measures in gene flow studies. Analysis of genetic surfaces could be readily applied to quantify the impact of ecological (e.g., temperature, moisture, natural barriers, elevation) and/or anthropogenic (e.g., roads, development, farming, forest management) variables on genetic structure. Understanding of the ecological and spatial behavior of a population is necessary for designing recovery plans, conserving corridors and understanding potential new population threats.

ACKNOWLEDGEMENTS

We would like to thank the following people for helpful comments and discussions: J Baumsteiger, R Dezzani, C Goldberg, J Kerby, M Morgan, A Robinson, S Spear, C Steele, T Rice, and three anonymous reviewers. In addition, we thank D Borcard and N Crookston for assistance in modifying the MODTTEST code, D Hall for Perl code to extract output values for analysis, and M Schwartz and C Pillgrim for work on the empirical black bear data set. Funding for this research was provided by: support under National Science Foundation award number DEB-0606458 to AS and MAM, King Fellowship (MAM), Washington State University Zoology Alumni Scholarship (MAM), EPA-STAR Fellowship (#FP-916695; MAM), and National Science Foundation award number DEB-0548415 to AS.

LITERATURE CITED

Arnaud, J.-F. 2003. Metapopulation genetic structure and migration pathways in the land snail *Heliz aspersa*: influence of landscape heterogeneity. - Landsc. Ecol. 18: 333-346.

Balloux, F. 2001. A computer program for the simulation of population genetics. - J. Hered. 92: 301-302.

Bockelmann, A. C., et al. 2003. Habitat differentiation vs. isolation-by-distance: the genetic population structure of *Elymus athericus* in European salt marshes. - Mol. Ecol. 12: 505-515.

Boots, B. and Getis, A. 1988. Point Pattern Analysis. - In: Thrall, G. I. (ed.) Scientific Geography Series. Sage Publications.

Bosserman, R. W. and Ragade, R. K. 1982. Ecosystem analysis using fuzzy set theory. - Ecological Modelling 16: 191-208.

Castellano, S. and Balletto, E. 2002. Is the partial mantel test inadequate? - Evolution 56: 1871-1873.

Corander, J., et al. 2003. Bayesian analysis of genetic differentiation between populations. - Genetics 163: 367-374.

Coulon, A., et al. 2004. Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. - Mol. Ecol. 13: 2841-2850.

Cressie, N. 1986. Kriging Nonstationary Data. - Journal of the American Statistical Association 81: 625-634.

Cressie, N. A. C. 1993. Statistics for spatial data. - Wiley.

Cushman, S. A., et al. 2006. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. - Am. Nat. 168: 486-499.

Dallas, J. F. 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. - Mammal Genome 5: 32-38.

Diggle, P. J. 2003. Statistical analysis of spatial point patterns. - Academic Press, New York.

Dutilleul, P. 1993. Modifying the t-test for assessing the correlation between two spatial processes. - Biometrics 49: 305-314.

Epperson, B. K. 2003. Geographical Genetics. - Princeton University Press.

Evanno, G., et al. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. - Mol. Ecol. 14: 2611-2620.

Fortin, M.-J. and Payette, S. 2002. How to test the significance of the relation between spatially autocorrelated data at the landscape scale: a case study using fire and forest maps. - Ecoscience 9: 213-218.

Fortin, M.-J. and Dale, M. 2005. Spatial analysis: a guide for ecologists. - Cambridge University Press.

Fotheringham, A. S., et al. 2002. Geographically weighted regression: the analysis of spatially varying relationships. - John Wiley and Sons, Ltd.

Gardner, R. H., et al. 1987. Neutral models for the analysis of broad-scale landscape pattern. - Landsc. Ecol. 1: 19-28.

Haining, R. 2003. Spatial Data Analysis: Theory and Practice. - Cambridge University Press.

Hedrick, P. 2005. 'Genetic restoration': a more comprehensive perspective than 'genetic rescue'. - Trends. Ecol. Evol. 20: 109.

Holderegger, R. and Wagner, H. 2006. A brief guide to Landscape Genetics. - Landsc. Ecol. 21: 793-796.

Holzhauer, S., et al. 2006. Effect of historic landscape change on the genetic structure of the bush-cricket *Metrioptera roeseli*. - Landsc. Ecol. 21: 891-899.

Isaaks, E. H. and Srivastava, R. M. 1989. An introduction to applied geostatistics. - Oxford University Press.

Keyghobadi, N., et al. 2005. Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassium smintheus*, vary with landscape connectivity. - Mol. Ecol. 14: 1897-1909.

Lam, N. 1983. Spatial interpolation methods: a review. - The American Cartographer 10: 129-149.

Lancaster, J. 2006. Using neutral landscapes to identify patterns of aggregation across resource points. - Ecography 29: 385-395.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? - Ecology 74: 1659-1673.

Legendre, P., et al. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. - Ecography 25: 601-615.

Lichstein, J. W., et al. 2002. Spatial autocorrelation and autoregressive models in ecology. - Ecol. Monogr. 72: 445-463.

McGarigal, K. and Cushman, S. 2005. The gradient concept of landscape structure. - In: Weins, J. and Moss, M. (eds.), Issues and perspectives in landscape ecology. Cambridge University Press, pp. 112-119.

McRae, B. H. 2006. Isolation by resistance. - Evolution 60: 1551-1561.

McRae, B. H. and Beier, P. 2007. Circuit theory predicts gene flow in plant and animal populations. - Proceedings of the National Academy of Sciences 104: 19885-19890.

Metternicht, G. I. 2003. Categorical fuzziness: a comparison between crisp and fuzzy class boundary modeling for mapping salt-affected soils using Landsat TM data and a classification based on anion ratios. - Ecological Modeling 168: pp. 371-389.

Mitasova, H. and Jaroslav, H. 1993. Interpolation by Regularized Spline with Tension: II. Application to Terrain Modeling and Surface Geometry Analysis. - Mathematical Geology 25: 657-669.

Mucina, L., et al. 1991. Trend Surface Analysis and Splines for Pattern Determination in Plant Communities. - In: Computer Assisted Vegetation Analysis. Kluwer Academic Publishers, pp. 355-371.

Pritchard, J. K., et al. 2000. Inference of population structure using multilocus genotype data. - Genetics 155: 945-959.

Raufaste, N. and Rousset, F. 2001. Are partial mantel tests adequate? - Evolution 55: 1703-1705.

Rempel, R. S. and Kushneriuk, R. S. 2003. The influence of sampling scheme and interpolation method on the power to detect spatial effects of forest birds in Ontario (Canada). - Landsc. Ecol. 18: 741-757.

Riley, S. P. D., et al. 2006. A southern California freeway is a physical and social barrier to gene flow in carnivores. - Mol. Ecol. 15: 1733-1741.

Rousset, F. 2002. Partial mantel tests: reply to Castellano and Balletto. - Evolution 56: 1874-1875.

Scribner, K., et al. 2001. Environmental correlates of toad abundance and population genetic diversity. - Biological Conservation 98: 201-210.

Shimatani, K. and Takahashi, M. 2003. On methods of spatial analysis for genotyped individuals. - Heredity 91: 173-180.

Shimatani, K. and Kubota, Y. 2004. Spatial analysis for continuously changing point patterns along a gradient and its application to an *Abies sachalinensis* population. - Ecological Modeling 180: 359-269.

Slatkin, M. and Arter, H. E. 1991. Spatial autocorrelation methods in population-genetics. - Am. Nat. 138: 499-517.

Sokal, R. R., et al. 1997. A simulation study of microevolutionary inferences by spatial autocorrelation analysis. - Biol. J. Linn. Soc. 60: 73-93.

Sokal, R. R., et al. 1998. Local spatial autocorrelation in biological variables. - Biol. J. Linn. Soc. 65: 41-62.

Spear, S. F., et al. 2005. Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). - Mol. Ecol. 14: 2553-2564.

Storfer, A., et al. 2007. Putting the 'landscape' in landscape genetics. - Heredity 98: 128-142.

Tobler, W. R. 1979. Smooth pycnophalactic interpolation for geographical regions. - Journal of the American Statistical Association 74: 519-530.

Wagner, H. H. and Fortin, M.-J. 2005. Spatial analysis of landscapes: concepts and statistics. - Ecology 86: 1975-1987.

Wand, M. P. and Jones, M. C. 1995. Kernal Smoothing. - Chapman & Hall.

Waples, R. S. and Gaggiotti, O. E. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. - Mol. Ecol. 15: 1419-1439.

Yfantis, E. A., et al. 1987. Efficiency of Kriging Estimation for Square, Triangular, and Hexagonal Grids. - Mathematical Geology 19: 183-205.

Zar, J. H. 1999. Biostatistical analysis. - Simon & Schuster.

TABLE

Table 1

Simulation parameters and power to detect landscape genetic structure with full dataset.

Simulation	Gen	M- B	M Within	Total N_e	Pops	MGS	Land	μR	Sig	SD
T1	0	0	Na	1000	2	None	A	0.071	0	0.027
T2	1	0	Na	1000	2	Patch	A	0.073	0	0.030
T3	5	0	Na	1000	2	Patch	A	0.542	5	0.225
T4	10	0	Na	1000	2	Patch	A	0.867	5	0.048
T5	50	0	Na	1000	2	Patch	A	0.986	5	0.003
T6	100	0	Na	1000	2	Patch	A	0.991	5	0.001
T7	250	0	Na	1000	2	Patch	A	0.991	5	0.001
T8	500	0	Na	1000	2	Patch	A	0.991	5	0.001
T9	1000	0	Na	1000	2	Patch	A	0.992	5	0.002
M1	500	0	Na	1000	2	Patch	A	0.991	5	0.001
M2	500	0.001	Na	1000	2	Patch	A	0.912	5	0.167
M3	500	0.005	Na	1000	2	Patch	A	0.990	5	0.002
M4	500	0.01	Na	1000	2	Patch	A	0.976	5	0.002
M5	500	0.05	Na	1000	2	Patch	A	0.949	5	0.014
M6	500	0.1	Na	1000	2	Patch	A	0.574	5	0.107
M7	500	0.2	Na	1000	2	Patch	A	0.175	1	0.120
HM1	500	0.001	0.05	1000	10	PXD	B	0.991	5	0.002
HM2	500	0.001	0.1	1000	10	PXD	B	0.991	5	0.002
HM3	500	0.001	0.2	1000	10	Patch	B	0.989	5	0.001
HM4	500	0.01	0.01	1000	10	Dist	B	0.866	4	0.123
HM5	500	0.01	0.05	1000	10	PXD	B	0.899	5	0.016
HM6	500	0.01	0.1	1000	10	PXD	B	0.976	5	0.012
HM7	500	0.01	0.2	1000	10	Patch	B	0.982	5	0.004
HM8	500	0.1	0.1	1000	10	Dist	B	0.897	5	0.020

Table 1 displays EASYPOP (Balloux 2001) simulation conditions as follows: simulation identification (Simulation), the number of generations populations post vicariance (Gen), probability of migration per individual per generation across the unsuitable habitat (Fig.1) (M Between), probability of migration between adjacent subpatches within populations (Fig.1) (M Within), total effective population size for the entire landscape (Total N_e), number of **simulated** (not genetically identified) populations or subpopulations (Pops), simulated model of genetic structure (MGS), and landscape used for simulation (Land). Simulated MGS was distance (dist), habitat patch, or patch-distance interaction (PXD). We held all other parameters constant

across simulations (see text). The table displays mean correlation coefficient (μ R), number significant correlations (Sig, X/5), and standard deviation (SD) for correlations between the model of landscape genetic structure and genetic surface. Simulation codes are as follows: T – time since vicariance, M- migration rate, and HM – hierarchical migration model. For population size (N) and unequal effective population size (UN), see Appendix 1b.

FIGURE LEGENDS

Figure 1 Flow chart of simulation and genetic surfacing methodology.

Step 1 – Define landscapes. Both landscapes consist of suitable and unsuitable habitat. Simulations in landscape A contain spatial dependency only, while simulations in landscape B contain autocorrelation or an interaction between spatial dependency and autocorrelation. **Step 2** – Simulate genetic data. Circles represent "populations" in EasyPop (labels correspond to habitat patches or subpatches on the corresponding landscape). Simulations in landscape B had two migrations rates labeled as M1 (between habitat patches) and M2 (migration rate within habitat patches). **Step 3** – subsample simulated data prior to analysis by sample size (n) and number of loci (l). **Step 4** – analysis in program STRUCTURE. **Step 5** – interpolate genetic surface. Patch and Distance on interpolated surfaces indicate condition under which data were simulated and n is the sample sized used to create that particular genetic surface. **Step 6** – model selection. We correlated each genetic surface with each model of landscape genetic structure (patch, distance, patch-distance interaction), tested significance, and selected model based on highest correlation coefficient. See methods for additional details for each step.

Figure 2 Empirical Landscape - North Idaho, USA.

The landscape was classified into habitat patches based on suitability as black bear habitat. Patch boundaries are superimposed on a shaded relief. Patches are labeled on the figure – 1,3 are suitable habitat while 2 is unsuitable habitat. Patch 4 was outside the study area and therefore excluded from the analysis. Points represent sample locations overlaid on the genetic surface and shaded relief. Light areas represent high ancestry values while dark areas represent low ancestry values.

Figure 3 Genetic Surfaces

Example genetic surfaces interpolated from STRUCTURE ancestry values including all samples and loci (see methods for details). Light areas represent high ancestry values while dark areas represent low ancestry values. A. No landscape genetic structure (T1: Complete admixture); B. Strong differentiation in relation to habitat patch (spatial dependency; T8: 500 generations post vicariance); C. With migration, genetic surfaces become more variable but are still differentiated (M4; 0.01 between patch migration probability); D. Internal patch structure results in a gradient of ancestry values across the landscape (HM4; 0.01 between patch and within patch migration probability).

Figure 4. Power of genetic surfacing with no subsampling.

a) Statistical power of the genetic surfacing approach to detect landscape genetic structure by sample size and number of loci; b) For all time since vicariance simulations (T), statistical power of the genetic surfacing technique to detect landscape genetic structure by time since vicariance and sample size, c) For 10 generations post vicariance (simulation T4), statistical power of the genetic surfacing technique to detect landscape genetic structure by sample size and number of loci, d) For all migration probability simulations (M), statistical power of the genetic surfacing technique to detect landscape genetic structure by probability of migration (migration rate) and sample size.

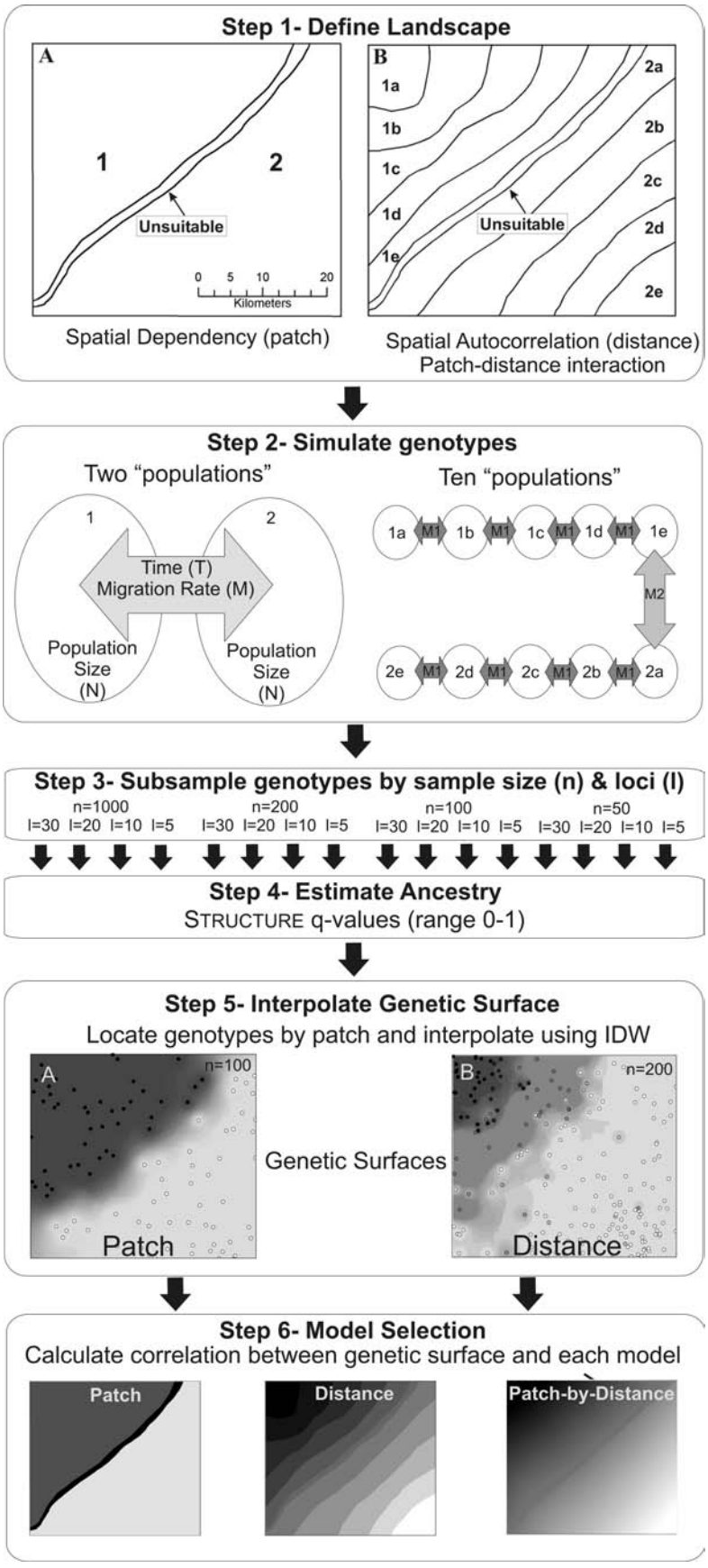
Figure 5 Accuracy of genetic surfacing.

In panels a – e, habitat patch model is shown in closed circles, distance model by open circles, and patch-distance interaction model by triangles. a) Overall proportion of correlations between genetic surface and model of genetic structure (habitat patch, distance, patch-distance interaction) significant by time since vicariance for each model. b) Overall proportion of correlations between genetic surface and model of genetic structure significant by migration probability for each model. c) Overall proportion of correlation between genetic surface and model of genetic structure significant by within patch migration probability. d) Proportion of analyses for which each model of genetic structure was selected for time since vicariance, e) Proportion of analyses for which each model of genetic structure was selected by migration probability. f) Proportion of analyses for which the "correct" model was selected by between and within migration probabilities.

Figure 6 Genetic Surface Significance by F_{ST}

Proportion of Pearson's correlation coefficients between the genetic surface and simulated landscape model significant (Y) by average F_{ST} value (X) across **all** analyses (simulation conditions, number of samples, number of loci). We can detect significant landscape process even when genetic structure is below the level generally considered biologically significant (<0.05).

Figure 1


Step 1- Define Landscape
A
1
2
Unsuitable
0 5 10 20
Kilometers
Spatial Dependency (patch)
B
1a
1b
1c
1d
1e
2a
2b
2c
2d
2e
Unsuitable
Spatial Autocorrelation (distance)
Patch-distance interaction
Step 2- Simulate genotypes
Two "populations"
Ten "populations"
Time (T)
Migration Rate (M)
Population Size (N)
Population Size (N)
M1
M2
Step 3- Subsample genotypes by sample size (n) & loci (l)
n=1000 l=30 l=20 l=10 l=5
n=200 l=30 l=20 l=10 l=5
n=100 l=30 l=20 l=10 l=5
n=50 l=30 l=20 l=10 l=5
Step 4- Estimate Ancestry
STRUCTURE q-values (range 0-1)
Step 5- Interpolate Genetic Surface
Locate genotypes by patch and interpolate using IDW
A
n=100
Patch
Genetic Surfaces
B
n=200
Distance
Step 6- Model Selection
Calculate correlation between genetic surface and each model
Patch
Distance
Patch-by-Distance

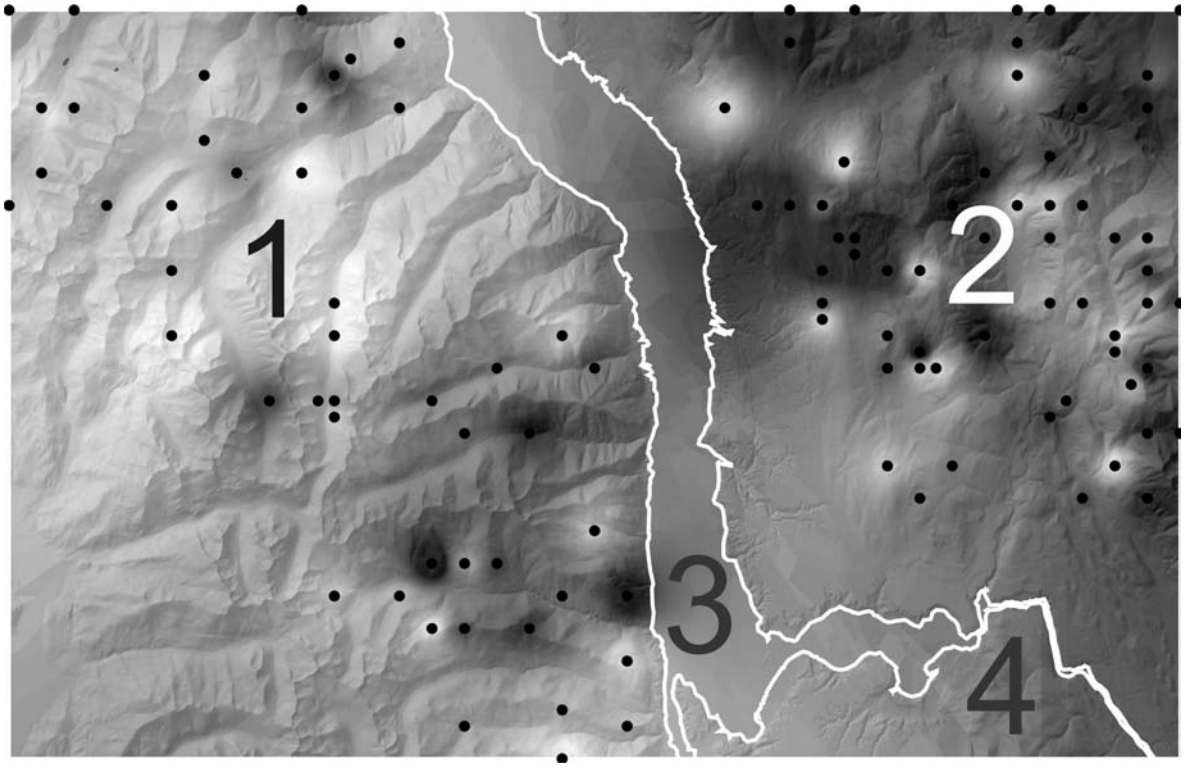


Figure 2

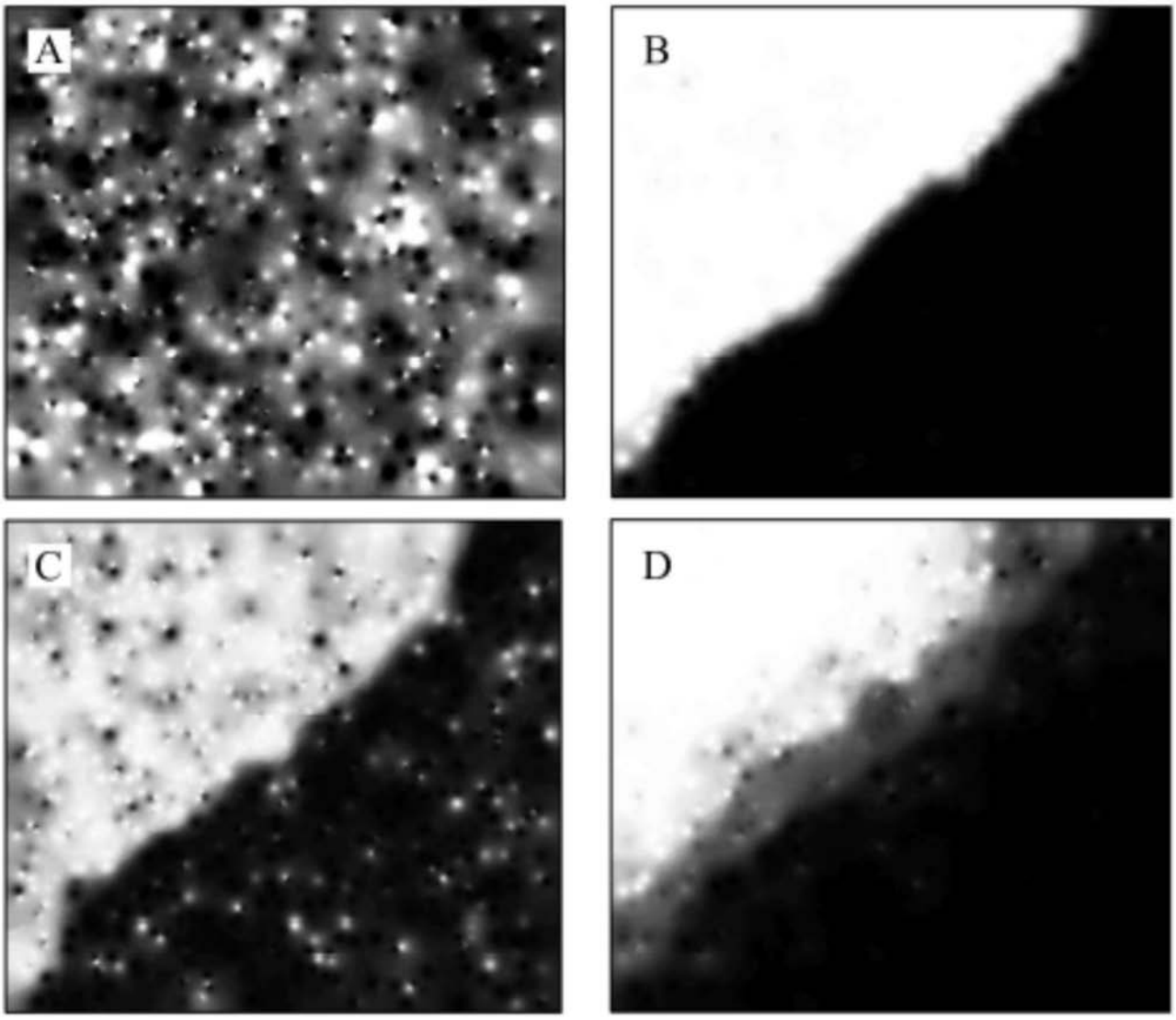


Figure 3

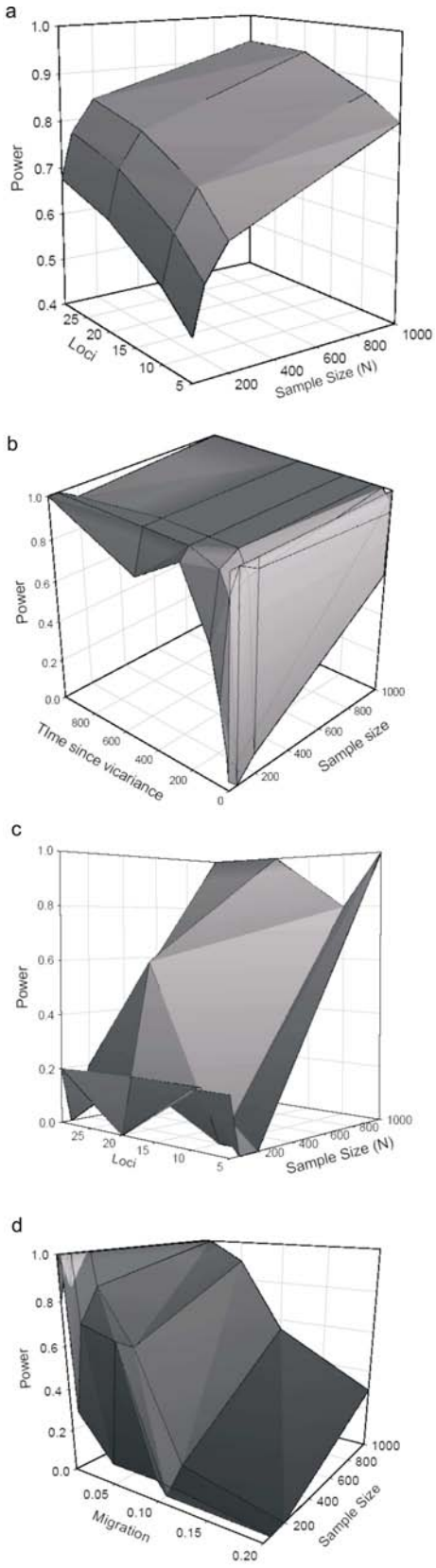

a
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Loci
25
20
15
10
5
Sample Size (N)
200
400
600
800
1000
b
Power
1.0
0.8
0.6
0.4
0.2
0.0
Time since vicariance
800
600
400
200
0
Sample size
200
400
600
800
1000
c
Power
1.0
0.8
0.6
0.4
0.2
0.0
Loci
25
20
15
10
5
Sample Size (N)
200
400
600
800
1000
d
Power
1.0
0.8
0.6
0.4
0.2
0.0
Migration
0.05
0.10
0.15
0.20
Sample Size
200
400
600
800
1000

Figure 4

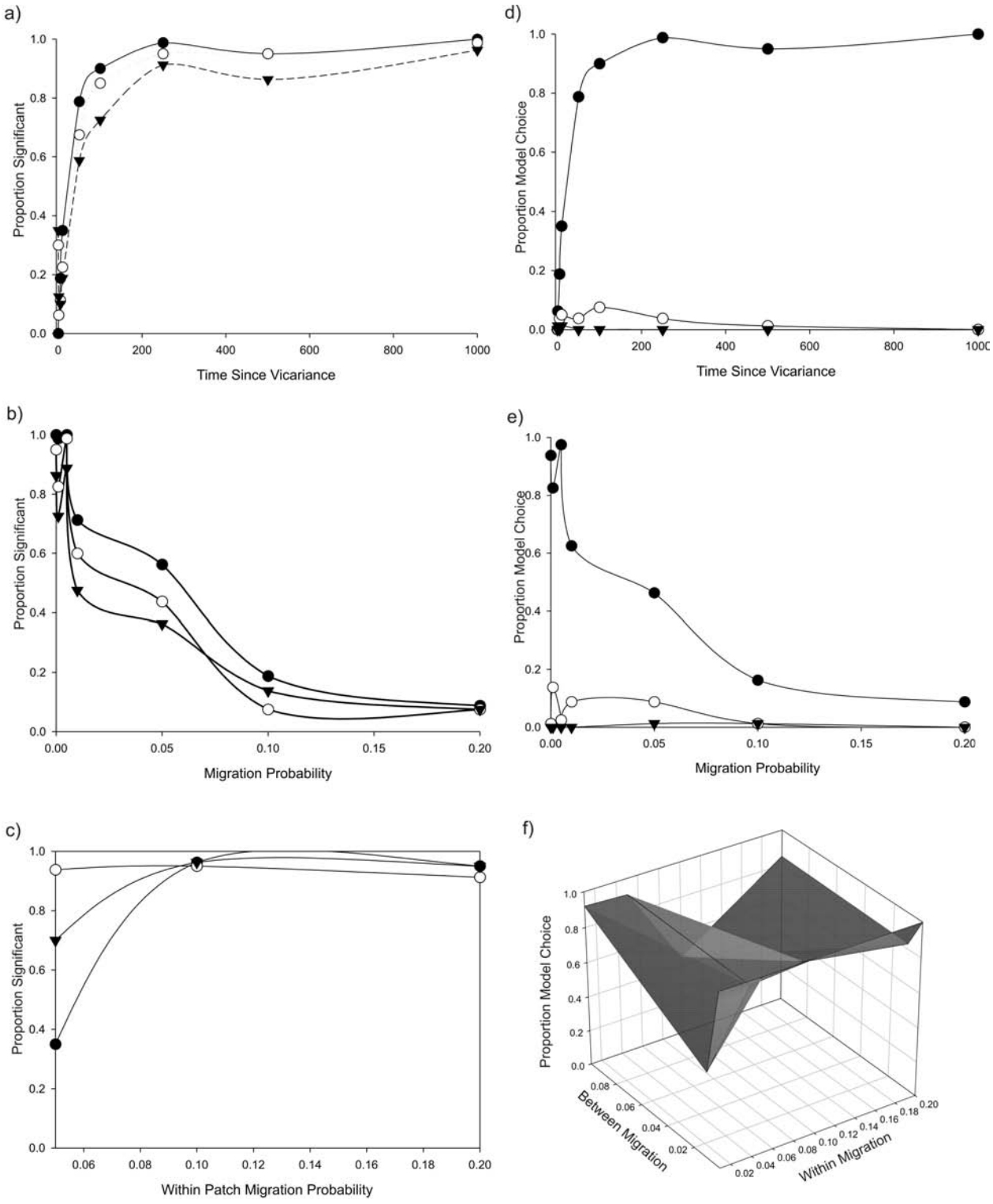

a)
Proportion Significant
Time Since Vicariance
d)
Proportion Model Choice
Time Since Vicariance
b)
Proportion Significant
Migration Probability
e)
Proportion Model Choice
Migration Probability
c)
Proportion Significant
Within Patch Migration Probability
f)
Proportion Model Choice
Between Migration
Within Migration

Figure 5

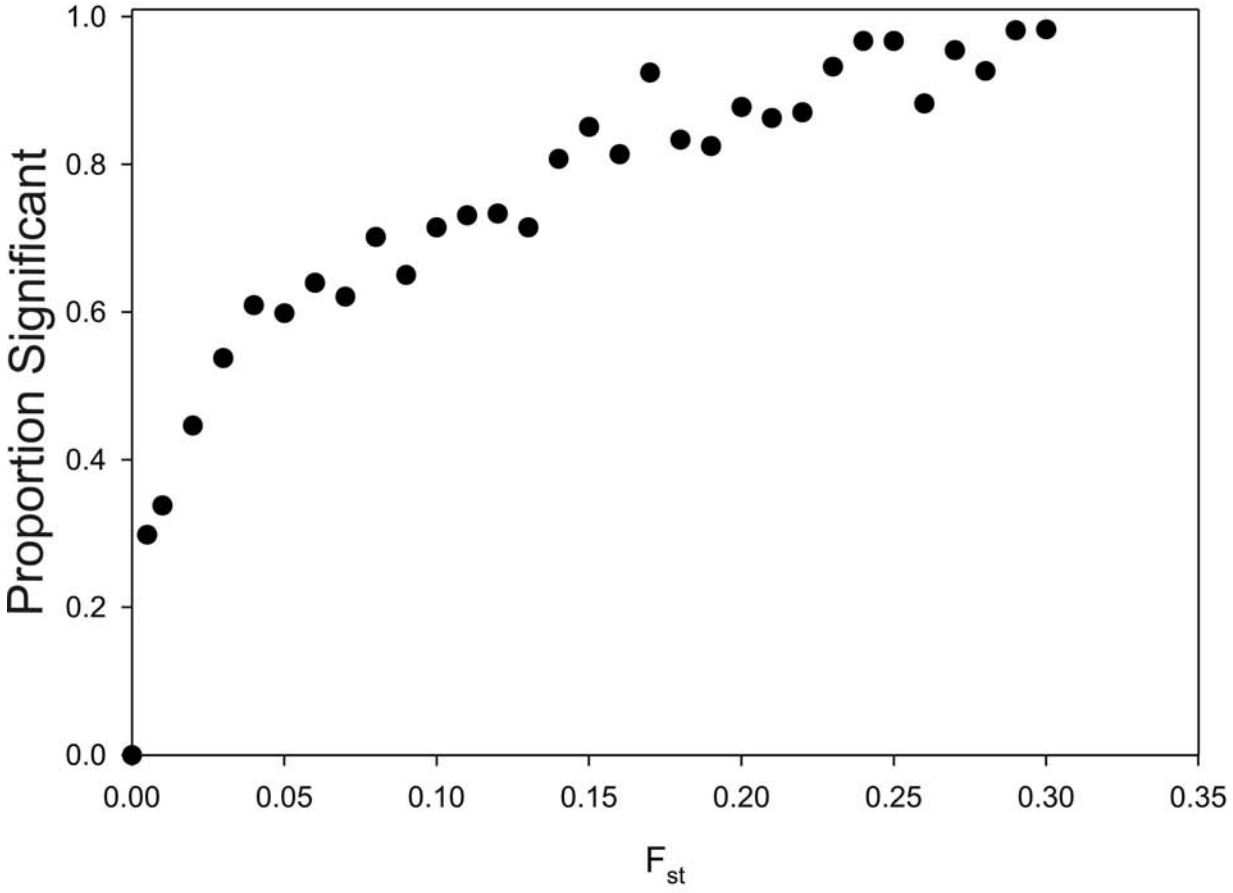

Proportion Significant
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
F st

Figure 6

APPENDICES

Appendix 1

This appendix includes a) additional details on parameter settings for utilized software and b) effects of effective samples size and proportional sampling on the power and accuracy of genetic surfacing.

Appendix 1.1. Additional information on parameters settings for utilized software.

EASYPOP: We simulated populations in EasyPop with the following universal parameter settings: diploid organisms, two sexes, equal sex ratio, random mating, free recombination between loci, mutation rate of 0.0005 (Dallas 1992), stepwise mutation with Kam events (0.05 probability), and 30 loci (see Table 1 for varied simulation conditions). All simulations began with each locus fixed at one allele and 5000 generations for burn-in under complete admixture; all populations reached equilibrium in fewer than 3000 generations.

STRUCTURE: STRUCTURE analysis was executed under the following conditions with no *a priori* definition of clusters or cluster number: 100,000 generation burn-in followed by 50,000 Markov Chain Monte Carlo (MCMC) iterations, equal alpha (starting value 1.0), uniform priors, correlated allele frequencies, admixture of populations, population number (*K*) 1-11, and lambda inferred from the data.

Appendix 1.2 Effect of unequal population sizes and proportional sampling

Not all extant populations are of equal size, therefore we simulated uneven effective populations with two migration probabilities (0.01 and 0): 500/500 (UN1, UN7, control), 600/400 (simulations UN2, UN8), 700/300 (simulations UN3, UN9), 800/200 (simulations UN4, UN10), 900/100 (simulations UN5, UN11) and 950/50 (simulations UN6, UN11) (Table A1). We subsampled these data at both absolute levels (n=200, 100, and 50) and proportion of the data (100%, 20%, 10%, and 5%). This was done to both provide a comparison among all simulation conditions, as well as to evaluate the relative power of detecting landscape genetic structure by sampling a percent of effective population size versus an absolute number of samples.

Overall small effective population size or unequal effective population size by suitable habitat patch had little impact on the power to detect significant landscape genetic structure (54/60 significant; Table A1). In the absence of migration, the genetic surfacing approach was extremely powerful (0.90 – 1) even with an effective population size of 100. In the presence of migration (0.01 migration probability), the genetic surfacing approach had >0.50 power at an effective population size as small as 100 (Appendix 2). In the absence of migration across the unsuitable habitat, all correlations were significant, while in the presence of low migration (probability = 0.01) landscape genetic structure was not detected when one effective population size was ≤ 100 or less (UN5, UN6; Table A1). However, statistical power to detect landscape genetic structure increases with equal numbers of samples per population relative to a random draw of genotypes from the entire landscape (Appendix 2).

Table A1

Simulation parameters and power to detect landscape genetic structure with full dataset.

Table A1 displays EASYPOP (Balloux 2001) simulation conditions as follows: simulation identification (Simulation), the number of generations populations post vicariance (Gen), probability of migration per individual per generation across the unsuitable habitat (Fig.1) (M Between), probability of migration between adjacent subpatches within populations (Fig.1) (M Within), total effective population size for the entire landscape (Total N_e), number of **simulated** (not genetically identified) populations or subpopulations (Pops), and effective population size per habitat patch (N_e/Patch). We held all other parameters constant across simulations (see main text). All simulations for effective population size and unequal population size were conducted in landscape A. The table displays number significant correlations (Sig, X/5), mean correlation coefficient (μ R), and standard deviation (SD) for correlations between the patch model of landscape genetic structure and genetic surface.

Simulation	Gen	M(B)	M(WI)	Total N_e	Pops	N_e/Patch	Sig	μR	SD
N1	500	0.01	Na	1000	2	500, 500	5	0.976	0.003
N2	500	0.01	Na	500	2	250, 250	5	0.944	0.010
N3	500	0.01	Na	200	2	100, 100	5	0.951	0.019
N4	500	0.01	Na	100	2	50, 50	5	0.928	0.012
N5	500	0.01	Na	50	2	25, 25	3	0.716	0.105
N6	500	0	Na	1000	2	500, 500	5	0.976	0.001
N7	500	0	Na	500	2	250, 250	5	0.985	0.003
N8	500	0	Na	200	2	100, 100	5	0.977	0.002
N9	500	0	Na	100	2	50, 50	5	0.968	0.003
N10	500	0	Na	50	2	25, 25	5	0.886	0.105
UN1	500	0.01	Na	1000	2	500, 500	5	0.976	0.003
UN2	500	0.01	Na	1000	2	600, 400	5	0.949	0.009
UN3	500	0.01	Na	1000	2	700, 300	5	0.911	0.025
UN4	500	0.01	Na	1000	2	800, 200	5	0.867	0.040
UN5	500	0.01	Na	1000	2	900, 100	3	0.458	0.376
UN6	500	0.01	Na	1000	2	950, 50	1	0.235	0.174
UN7	500	0	Na	1000	2	500, 500	5	0.991	0.001
UN8	500	0	Na	1000	2	600, 400	5	0.989	0.002
UN9	500	0	Na	1000	2	700, 300	5	0.985	0.003
UN10	500	0	Na	1000	2	800, 200	5	0.971	0.009
UN11	500	0	Na	1000	2	900, 100	5	0.928	0.033
UN12	500	0	Na	1000	2	950, 50	5	0.878	0.012

Appendix 2 Extended table of results by sample size and number of loci.

Simulation data summarized by sample size (N, on left) and number of loci (Loci, on right). Headings are as follows: Simulation (Sim): Simulation (conditions defined in Table 1, Appendix 1b), Model: landscape model used to simulate data (habitat patch (Patch), distance (Dist), patch-distance interaction (PXD); see text for definitions), Power (P): statistical power (1-β (Type II error)) to detect landscape genetic structure with the genetic surface, Model Choice (MC): the proportion of times the correct landscape model was selected out of the total number of significant simulations, F_{ST}: average F_{ST} between population clusters and A (allelic divergence): average allelic divergence between population clusters as estimated in STRUCTURE.

Sim	Model	N	P	MC	Fst	A	Loci	P	MC	Fst	A
T1	Patch	1000	na	na	0.073	0.081	30	na	na	0.069	0.085
(0 gen)	Patch	200	na	na	0.042	0.040	20	na	na	0.055	0.064
	Patch	100	na	na	0.064	0.085	10	na	na	0.062	0.086
	Patch	50	na	na	0.040	0.043	5	na	na	0.033	0.015
T2	Patch	1000	na	na	0.103	0.110	30	na	na	0.091	0.130
(1 gen)	Patch	200	na	na	0.064	0.081	20	na	na	0.087	0.088
	Patch	100	na	na	0.056	0.067	10	na	na	0.051	0.047
	Patch	50	na	na	0.059	0.063	5	na	na	0.053	0.057
T3	Patch	1000	0.60	0.80	0.069	0.111	30	0.30	1.00	0.047	0.071
(5 gen)	Patch	200	0.10	1.00	0.043	0.045	20	0.20	0.80	0.056	0.092
	Patch	100	0.00	na	0.054	0.059	10	0.15	0.75	0.047	0.055
	Patch	50	0.05	1.00	0.025	0.020	5	0.10	0.67	0.041	0.018
T4	Patch	1000	0.95	0.86	0.064	0.095	30	0.35	1.00	0.058	0.083
(10 gen)	Patch	200	0.20	0.80	0.047	0.061	20	0.45	0.82	0.053	0.070
	Patch	100	0.10	0.67	0.037	0.035	10	0.30	0.86	0.043	0.053
	Patch	50	0.15	1.00	0.047	0.049	5	0.30	0.75	0.040	0.032

T5	Patch	1000	1.00	1.00	0.093	0.238	30	1.00	1.00	0.076	0.188
(50 gen)	Patch	200	0.90	0.95	0.075	0.183	20	0.85	1.00	0.068	0.166
	Patch	100	0.85	0.94	0.067	0.138	10	0.75	1.00	0.076	0.167
	Patch	50	0.40	0.89	0.046	0.057	5	0.55	0.79	0.062	0.094
T6	Patch	1000	1.00	1.00	0.147	0.411	30	0.95	1.00	0.113	0.294
(100 gen)	Patch	200	1.00	0.91	0.121	0.342	20	1.00	1.00	0.118	0.304
	Patch	100	0.95	0.90	0.104	0.270	10	0.90	0.90	0.114	0.310
	Patch	50	0.65	0.87	0.077	0.153	5	0.75	0.79	0.103	0.268
T7	Patch	1000	1.00	1.00	0.236	1.087	30	1.00	0.95	0.199	0.787
(250 gen)	Patch	200	1.00	1.00	0.205	0.863	20	0.95	1.00	0.198	0.823
	Patch	100	1.00	1.00	0.196	0.780	10	1.00	1.00	0.211	0.875
	Patch	50	0.95	0.86	0.159	0.572	5	1.00	0.91	0.188	0.817
T8	Patch	1000	1.00	1.00	0.311	1.608	30	0.95	0.95	0.275	1.172
(500 gen)	Patch	200	1.00	1.00	0.277	1.320	20	0.95	1.00	0.273	1.228
	Patch	100	1.00	1.00	0.268	1.197	10	0.95	1.00	0.277	1.382
	Patch	50	0.80	0.94	0.238	0.954	5	0.95	1.00	0.270	1.308
T9	Patch	1000	1.00	1.00	0.387	2.697	30	1.00	1.00	0.359	2.112
(1000 gen)	Patch	200	1.00	1.00	0.355	2.204	20	1.00	1.00	0.360	2.132
	Patch	100	1.00	1.00	0.341	1.972	10	1.00	1.00	0.349	2.111
	Patch	50	1.00	1.00	0.324	1.676	5	1.00	1.00	0.340	2.190
M1	Patch	1000	1	1.00	0.311	1.608	30	1	0.95	0.275	1.172
(0 mp)	Patch	200	1	1.00	0.277	1.320	20	1	1.00	0.273	1.228
	Patch	100	1	1.00	0.268	1.197	10	1	1.00	0.277	1.382
	Patch	50	1	0.94	0.238	0.954	5	1	1.00	0.270	1.308
M2	Patch	1000	1	0.80	0.213	0.912	30	1	0.85	0.195	0.733
(0.001 mp)	Patch	200	1	0.90	0.195	0.764	20	1	0.90	0.197	0.732
	Patch	100	1	1.00	0.185	0.689	10	1	0.89	0.191	0.757
	Patch	50	1	0.75	0.178	0.627	5	0.95	0.78	0.188	0.769

M3	Patch	1000	1	1.00	0.306	1.534	30	1	1.00	0.271	1.162
(0.005 mp)	Patch	200	1	1.00	0.280	1.300	20	1	0.95	0.281	1.307
	Patch	100	1	1.00	0.263	1.134	10	1	1.00	0.267	1.148
	Patch	50	1	0.90	0.248	0.973	5	1	0.95	0.278	1.325
M4	Patch	1000	1	0.90	0.077	0.205	30	0.9	1.00	0.063	0.151
(0.01 mp)	Patch	200	0.9	0.88	0.070	0.169	20	0.9	0.94	0.068	0.158
	Patch	100	0.7	0.79	0.057	0.101	10	0.65	0.77	0.062	0.123
	Patch	50	0.3	1.00	0.046	0.053	5	0.4	0.63	0.058	0.097
M5	Patch	1000	0.9	0.83	0.057	0.129	30	0.8	1.00	0.043	0.084
(0.05 mp)	Patch	200	0.6	0.92	0.048	0.070	20	0.8	0.88	0.062	0.125
	Patch	100	0.7	0.77	0.061	0.083	10	0.4	0.88	0.058	0.084
	Patch	50	0.1	0.50	0.052	0.065	5	0.25	0.00	0.056	0.054
M6	Patch	1000	0.6	0.82	0.062	0.083	30	0.3	0.83	0.075	0.112
(0.1 mp)	Patch	200	0.1	1.00	0.051	0.053	20	0.2	0.75	0.057	0.086
	Patch	100	0	na	0.062	0.066	10	0.1	na	0.056	0.029
	Patch	50	0.1	1.00	0.045	0.044	5	0.15	1.00	0.032	0.019
M7	Patch	1000	na	na	0.045	0.122	30	na	na	0.060	0.078
(0.2 mp)	Patch	200	na	na	0.044	0.063	20	na	na	0.056	0.088
	Patch	100	na	na	0.042	0.035	10	na	na	0.049	0.055
	Patch	50	na	na	0.028	0.029	5	na	na	0.044	0.028
N1	Patch	500	0.90	0.90	0.311	1.608	30	0.90	0.63	0.275	1.172
1000	Patch	200	0.90	0.88	0.277	1.320	20	0.90	0.77	0.273	1.228
	Patch	100	0.65	0.79	0.268	1.197	10	0.65	0.94	0.277	1.382
	Patch	50	0.40	1.00	0.238	0.954	5	0.40	1.00	0.270	1.308
N2	Patch	500	0.72	0.95	0.084	0.133	30	0.72	0.90	0.052	0.088
500	Patch	200	0.64	0.88	0.087	0.133	20	0.64	0.86	0.078	0.080
	Patch	100	0.56	1.00	0.077	0.087	10	0.56	0.88	0.066	0.089
	Patch	50	0.40	0.80	0.063	0.041	5	0.40	1.00	0.065	0.057
	Patch	10	0.70	1.00	0.015	0.001					

N3	Patch	200	0.73	0.84	0.155	0.169	30	0.70	0.56	0.052	0.120
200	Patch	100	0.47	0.88	0.134	0.151	20	0.73	0.86	0.091	0.119
	Patch	50	0.30	0.77	0.116	0.108	10	0.47	0.86	0.131	0.088
	Patch	40	0.52	0.82	0.107	0.108	5	0.30	0.86	0.133	0.054
	Patch	20	0.52	0.80	0.075	0.034					
	Patch	10	0.40	0.50	0.024	0.000					
N4	Patch	100	0.08	0.93	0.162	0.117	30	0.52	na	0.028	0.079
100	Patch	50	0.50	0.86	0.155	0.095	20	0.52	0.80	0.093	0.057
	Patch	20	0.30	0.67	0.064	0.024	10	0.40	0.85	0.106	0.053
	Patch	10	0.30	0.00	0.041	0.001	5	0.08	1.00	0.132	0.000
N5	Patch	40	0.00	1.00	0.214	0.146	30	0.50	na	0.047	0.053
50	Patch	8	1.00	0.33	0.067	0.019	20	0.30	0.67	0.060	0.058
	Patch						10	0.30	1.00	0.109	0.028
	Patch						5	0.00	0.80	0.118	0.027
N6	Patch	1000	1.00	1.00	0.311	0.205	30	1.00	0.95	0.063	0.151
1000	Patch	200	1.00	1.00	0.277	0.169	20	1.00	1.00	0.068	0.158
	Patch	100	0.95	1.00	0.268	0.101	10	1.00	1.00	0.062	0.123
	Patch	50	0.73	0.94	0.238	0.053	5	0.95	1.00	0.058	0.097
N7	Patch	500	0.83	1.00	0.566	5.137	30	0.73	1.00	0.490	3.980
500	Patch	200	0.83	1.00	0.545	4.540	20	0.83	0.96	0.538	4.014
	Patch	100	0.83	1.00	0.537	3.964	10	0.83	1.00	0.550	3.919
	Patch	50	0.93	1.00	0.519	3.424	5	0.83	0.95	0.555	3.643
	Patch	20	0.77	0.88	0.511	2.719					
N8	Patch	200	0.87	1.00	0.690	5.571	30	0.93	0.85	0.559	4.200
200	Patch	100	0.90	1.00	0.684	5.046	20	0.77	0.73	0.658	4.418
	Patch	50	0.68	0.95	0.671	4.436	10	0.87	0.91	0.703	4.120
	Patch	40	0.60	0.90	0.672	4.261	5	0.90	0.86	0.700	3.963
	Patch	20	0.68	0.60	0.655	3.580					
	Patch	10	0.64	0.00	0.557	2.427					

N9	Patch	100	0.60	1.00	0.767	5.574	30	0.68	0.94	0.580	4.063
100	Patch	50	0.60	1.00	0.752	4.898	20	0.60	0.88	0.705	4.087
	Patch	20	0.90	0.88	0.733	4.012	10	0.68	0.87	0.751	3.890
	Patch	10	0.50	0.57	0.657	2.832	5	0.64	0.94	0.753	3.431
N10	Patch	40	0.50	0.60	0.817	4.638	30	0.60	0.60	0.390	2.745
50	Patch	8	0.90	0.00	0.730	2.850	20	0.60	0.33	0.562	2.536
	Patch						10	0.90	0.50	0.570	2.196
	Patch						5	0.50	0.50	0.612	1.621
UN1	Patch	1000	1	0.90	0.077	0.205	30	0.9	1.00	0.063	0.151
500	Patch	200	0.9	0.88	0.070	0.169	20	0.9	0.94	0.068	0.158
	Patch	100	0.7	0.79	0.057	0.101	10	0.65	0.77	0.062	0.123
	Patch	50	0.3	1.00	0.046	0.053	5	0.4	0.63	0.058	0.097
UN2	Patch	1000	0.9	1.00	0.038	0.136	30	0.75	0.93	0.051	0.094
400	Patch	200	0.7	0.93	0.057	0.088	20	0.65	1.00	0.044	0.112
	Patch	100	0.5	1.00	0.056	0.083	10	0.5	1.00	0.057	0.059
	Patch	50	0.1	1.00	0.056	0.026	5	0.25	1.00	0.055	0.067
UN3	Patch	1000	0.9	1.00	0.066	0.116	30	0.6	1.00	0.061	0.080
300	Patch	200	0.6	1.00	0.062	0.068	20	0.45	1.00	0.068	0.123
	Patch	100	0.2	0.80	0.062	0.066	10	0.4	1.00	0.076	0.072
	Patch	50	0	Na	0.068	0.079	5	0.15	0.67	0.055	0.055
UN4	Patch	1000	0.9	1.00	0.039	0.107	30	0.5	1.00	0.046	0.110
200	Patch	200	0.4	0.88	0.062	0.050	20	0.35	0.86	0.058	0.076
	Patch	100	0.2	1.00	0.054	0.074	10	0.3	1.00	0.048	0.053
	Patch	50	0.1	0.00	0.063	0.049	5	0.25	0.80	0.067	0.041
UN5	Patch	1000	0.6	1.00	0.047	0.090	30	0.2	1.00	0.060	0.137
100	Patch	200	0.2	0.80	0.052	0.097	20	0.3	0.83	0.040	0.065
	Patch	100	0.1	na	0.063	0.067	10	0.1	1.00	0.049	0.033
	Patch	50	0	na	0.073	0.058	5	0.2	1.00	0.087	0.076

UN6	Patch	1000	0.2	1.00	0.033	0.063	30	0.05	1.00	0.042	0.064
50	Patch	200	0	na	0.056	0.058	20	0.05	1.00	0.045	0.086
	Patch	100	0	na	0.049	0.067	10	0.05	1.00	0.077	0.037
	Patch	50	0	na	0.083	0.021	5	0.05	1.00	0.058	0.024
UN7	Patch	1000	1	1.00	0.311	1.608	30	0.95	0.95	0.275	1.172
500	Patch	200	1	1.00	0.277	1.320	20	0.95	1.00	0.273	1.228
	Patch	100	1	1.00	0.268	1.197	10	0.95	1.00	0.277	1.382
	Patch	50	0.8	0.94	0.238	0.954	5	0.95	1.00	0.270	1.308
UN8	Patch	1000	1	0.91	0.268	1.770	30	1	0.95	0.282	1.415
400	Patch	200	1	1.00	0.282	1.455	20	1	1.00	0.295	1.446
	Patch	100	1	1.00	0.298	1.293	10	1	1.00	0.305	1.409
	Patch	50	1	1.00	0.331	1.130	5	1	0.95	0.297	1.378
UN9	Patch	1000	1	0.80	0.254	1.779	30	0.95	1.00	0.270	1.320
300	Patch	200	1	1.00	0.273	1.423	20	1	0.90	0.289	1.355
	Patch	100	1	1.00	0.287	1.263	10	1	0.95	0.287	1.472
	Patch	50	1	1.00	0.318	1.059	5	1	0.90	0.286	1.375
UN10	Patch	1000	1	0.83	0.277	1.961	30	1	0.85	0.275	1.623
200	Patch	200	1	0.80	0.305	1.541	20	1	0.95	0.326	1.613
	Patch	100	1	1.00	0.326	1.315	10	0.9	0.83	0.331	1.537
	Patch	50	0.8	1.00	0.359	1.049	5	0.9	0.89	0.336	1.093
UN11	Patch	1000	1	0.91	0.257	2.580	30	0.75	0.47	0.307	1.805
100	Patch	200	1	0.44	0.309	2.054	20	0.65	0.62	0.343	1.642
	Patch	100	0.6	0.15	0.374	1.492	10	0.7	0.57	0.326	1.752
	Patch	50	0.2	1.00	0.402	1.001	5	0.65	0.69	0.366	1.927
UN12	Patch	1000	1	0.95	0.127	1.426	30	0.5	0.70	0.094	1.207
50	Patch	200	0.5	0.67	0.197	0.855	20	0.5	0.60	0.188	1.087
	Patch	100	0.3	0.42	0.215	0.688	10	0.5	0.90	0.260	0.736
	Patch	50	0.3	1.00	0.288	0.370	5	0.55	0.91	0.285	0.309

HM1	PXD	1000	1	1.00	0.264	1.225	30	1	1.00	0.245	1.033
0.05	PXD	200	1	1.00	0.242	1.054	20	1	1.00	0.242	1.041
0.001	PXD	100	1	1.00	0.225	0.923	10	0.95	1.00	0.248	1.107
	PXD	50	0.9	0.95	0.203	0.767	5	0.9	0.95	0.200	0.790
HM2	PXD	1000	1	1.00	0.270	1.380	30	1	0.95	0.250	1.146
0.1	PXD	200	1	1.00	0.246	1.185	20	1	0.95	0.254	1.158
0.001	PXD	100	1	0.95	0.240	1.084	10	0.95	1.00	0.238	1.148
	PXD	50	0.9	0.89	0.225	0.907	5	0.95	0.95	0.238	1.104
HM3	Patch	1000	0.8	1.00	0.328	2.171	30	0.8	0.94	0.296	1.718
0.2	Patch	200	0.8	1.00	0.306	1.851	20	0.8	1.00	0.296	1.689
0.001	Patch	100	0.8	1.00	0.300	1.721	10	0.8	1.00	0.323	1.963
	Patch	50	0.8	0.93	0.274	1.423	5	0.75	1.00	0.293	1.795
HM4	Dist	1000	0.8	0.42	0.268	2.034	30	0.8	0.43	0.264	1.693
0.01	Dist	200	0.8	0.25	0.247	1.594	20	0.75	0.38	0.249	1.684
0.01	Dist	100	0.8	0.33	0.235	1.435	10	0.75	0.33	0.231	1.435
	Dist	50	0.8	0.50	0.210	1.138	5	0.8	0.30	0.216	1.390
HM5	PXD	1000	0.8	1.00	0.179	0.625	30	0.75	0.92	0.150	0.491
0.05	PXD	200	0.8	1.00	0.159	0.545	20	0.8	0.88	0.154	0.514
0.01	PXD	100	0.7	0.67	0.144	0.449	10	0.65	0.80	0.158	0.516
	PXD	50	0.6	0.88	0.131	0.366	5	0.6	1.00	0.152	0.464
HM6	PXD	1000	1	1.00	0.164	0.652	30	1	1.00	0.138	0.530
0.1	PXD	200	1	0.85	0.141	0.555	20	1	0.85	0.140	0.507
0.01	PXD	100	1	0.70	0.135	0.492	10	0.9	0.74	0.140	0.505
	PXD	50	0.9	0.76	0.118	0.393	5	0.95	0.72	0.140	0.549
HM7	Patch	1000	1	1.00	0.161	0.545	30	1	0.90	0.131	0.450
0.2	Patch	200	1	1.00	0.143	0.472	20	1	1.00	0.130	0.423
0.01	Patch	100	1	0.79	0.129	0.404	10	0.95	0.95	0.151	0.450
	Patch	50	0.9	0.76	0.124	0.359	5	0.85	0.71	0.143	0.458

HM8	Dist	1000	0.9	0.47	0.076	0.197	30	0.95	0.53	0.057	0.130
0.1	Dist	200	0.8	0.54	0.075	0.158	20	0.85	0.50	0.087	0.186
0.1	Dist	100	0.6	0.17	0.062	0.103	10	0.5	0.18	0.061	0.116
	Dist	50	0.5	0.25	0.061	0.069	5	0.3	0.50	0.069	0.096

CHAPTER 2 - Frog and Toad (not) Together: Quantifying differences in species niche and distribution

Melanie A. Murphy[1], Jeffrey S. Evans[2], Charles Peterson[3], Andrew Storfer[1]

[1] School of Biological Sciences, Department of Biology, Washington State University, Pullman WA, 99164, USA [2] USDA Forest Service, Rocky Mountain Research Station, 1221 S. Main, Moscow, ID 83483, USA [3] Department of Biology, Idaho State University, Pocatello, ID 82309

Corresponding Author:

Melanie A. Murphy

School of Biological Sciences

Washington State University

Pullman WA, 99164

208-883-1069, (mamurphy@wsu.edu)

Running title: Predicting anuran breeding niche

For submission to Ecology

ABSTRACT

A fundamental goal in ecology and biogeography is to understand processes driving differences in species' distributions. Divergent fine-scale distributions could arise from random spatial process, differences in species' suitable environmental conditions ("Grinellian" niche theory), or restricted occupancy of suitable environment due to limited dispersal ("dispersal limited" niche theory). These concepts provide a theoretical background for predicting spatially explicit species distribution models. The *Pseudacris maculata* - *Bufo boreas* system in Yellowstone National Park is well-suited to test alternative processes explaining observed differences in distribution. Both anuran species breed in similar habitats (shallow, temporary wetlands) but have differing dispersal capabilities and only rarely co-occur. We found the distributions are significantly different (p-value = 0.032), with *P. maculata* occupying more sites than *B. boreas*. We were able to separate species niche base on environmental conditions with models developed in one species unable to reciprocally predict the other species (percent correctly classified (PCC) < 5%). The breeding niche of *P. maculata* in Yellowstone was defined by temperature, abiotic wetland characteristics, rainfall, dominant cover (shrub), and absence of thermal influence (percent correctly classified (PCC) 98%, out of bag (OOB) error 21.1 %,). The breeding niche of *B. boreas* in Yellowstone was defined by abiotic wetland characteristics, topographic complexity, rainfall, dominant cover and temperature (PCC 95.8%, OOB error 23.9%). The dispersal limited niche theory best explained the observed distribution for both species. In these models, distance to nearest source population was the most important predictor for both species. This suggests that the availability of a source population is extremely important for site occupancy and reinforcing the importance of metapopulation dynamics. Finally, niche-based distribution

models from our study can be used for targeted sampling of rare species and to predict distributional shifts under climate change scenarios.

Keywords: niche theory, species distribution model, topographic variables, algorithmic models, Grinellian niche theory, dispersal limited niche theory, Boreal chorus frog, Boreal toad, amphibians

INTRODUCTION

A fundamental goal in ecology and biogeography is to understand processes driving differences in species' distributions (Pulliam 2000, Guisan and Thuiller 2005). In a realized environment (Ackerly 2003), divergent fine-scale distributions could result from random spatial process, differences in suitable environmental conditions ("Grinellian" niche theory), or restricted occupancy of the suitable environment due to limited dispersal ("dispersal limited" niche theory) (Pulliam 2000). These concepts provide a theoretical background for predicting spatially explicit species distribution models (Guisan and Zimmermann 2000).

The applicable niche theory may depend on the ecological questions and study organism(s), with respective predictions of processes driving distribution divergence. At fine-scales, all available habitat may satisfy species' environmental requirements and observed differences in distributions may be explained by random spatial process (Alonso et al. 2006). Alternatively, species presence may be explained by environmental conditions (Pulliam 2000) which can be used to predict suitable habitat (Guisan and Thuiller 2005). However, otherwise suitable habitat may be unoccupied due to species dispersal constraints or intervening hostile matrix. If the underlying research goal is to predict potential habitat (Rehfeldt 2006), this is not problematic. However, if the goal is to predict the realized distribution, predictions based solely on environmental conditions may result in an over-prediction of species distribution. Distribution

models based on the dispersal limited niche theory incorporate both environmental condition and geographic accessibility (Pulliam 2000).

Quantifying species' distribution based on environmental requirements is crucial for amphibian ecology and conservation (Semlitsch 2002). Amphibians are declining globally (~33% of species threatened) with habitat as the leading causal factor even in unexplained declines (Storfer 2003, Stuart et al. 2004, Gardner et al. 2007). Distribution models based on niche theory can be used in identification of reserve priorities (Garcia 2006, Pawar et al. 2007) and predict shifts in suitable environmental conditions due to climate change (Araujo et al. 2006). Several environmental conditions limit amphibian occurrence including: physiological constraints (Bartelt and Peterson 2005), hydroperiod (Werner et al. 2007b), cover requirements (Babbit et al. 2006), and disturbance (Arens et al. 2007, Hossack and Corn 2007) . In addition, amphibians generally have limited dispersal abilities (Marsh et al. 1999, Arens et al. 2007), may operate as metapopulations (Smith and Green 2005), and geographic features can act as barriers to dispersal (Lougheed et al. 1999, Funk et al. 2005). For these reasons, observed distributions may be explained by dispersal limited niche theory.

The *Pseudacris maculata* (boreal chorus frog) - *Bufo boreas* (boreal toad) system in Yellowstone National Park is well-suited to test alternative processes explaining observed differences in distribution. *Pseudacris maculata* and *B. boreas* share several basic breeding habitat requirements including semi-permanent wetlands and warm shallow waters for rearing (Koch and Peterson 1995). However, the observed breeding distributions of these species in Yellowstone appear divergent: *P. maculata* is relatively widespread, while *B. boreas* is locally abundant but patchily distributed and thought to be in decline (Carey 1993, Koch and Peterson 1995, Muths et al. 2003).

We investigated differences in fine-scale species distribution of *P. maculata* and *B. boreas* to address three central questions: i) are observed distributions of *P. maculata* and *B. boreas* significantly different?, ii) can *P. maculata* and *B. boreas* niche be separated based on environmental requirements (Grinellian niche)? and iii) are the observed distributions best explained by a Grinellian or dispersal limited niche theory? While upland habitat is important for amphibian persistence (Semlitsch 2002, Denoel and Lehmann 2006), niche is defined from empirical observation of successful reproduction in an environment (Guisan and Thuiller 2005). Therefore, we predict breeding niche and distribution for these two species. In answering these questions, we developed modelling and evaluation methodology addressing current limitations of niche-based species distribution models (Guisan and Thuiller 2005, Araujo and Guisan 2006).

METHODS

Field sampling

To capture the environmental range of variability in observations, we stratified suitable wetlands from the National Wetlands Inventory (NWI) (USFWS 2001) (Munger et al. 1998) for Yellowstone National Park by elevation and precipitation excluding closed or inaccessible areas (Fig. 7). Elevation impacts time to metamorphosis and may limit gene flow among populations (Palo et al. 2003, Giordano et al. 2007). Precipitation impacts wetland persistence (Corn 2003), amphibian breeding behavior (Bosch and Martinez-Solano 2003), and amphibian genetic connectivity (Murphy et al. in review). We randomized wetlands within strata and sampled at least twenty sites per stratum. Because occurrences of *B. boreas* in Yellowstone are rare (~3% of wetlands; D. Patla pub. comm.), known *B. boreas* breeding sites were visited in addition to the stratification to ensure adequate sampling (Fig. 7). Geographic Positioning System (GPS) coordinates were taken at each site with a Garmin 12XL recreation-grade device with at least 50

points averaged per location. We detected individuals (adults, metamorphs, larvae) by walking the parameter of the site and conducting random dip-net sweeps (D. Patla pub. comm.). We recorded presence of adult breeding activity (amplexus, calling), eggs, larvae, and metamorphs. Abundance was ranked as high (more than 100 larvae or 10 breeding adults), medium (at least 30 larvae or 3 breeding adults), low (less than 30 larvae or 1 breeding adults), and none (no breeding activity detected). We derived probability of detection using standard protocols (MacKenzie et al. 2002) from ongoing monitoring efforts concurrent with our field sampling (*P. maculata* – 0.89, *B. boreas* – 0.88; D. Patla, pub. comm.).

Spatial distribution

We described the spatial distribution for each species by evaluating: clustering in species presence (Batcheler 1971), autocorrelation in species abundance (Moran 1950, Geary 1954), and range of autocorrelation (Cliff and Haggett 1988). We estimated clustering using nearest-neighbor statistic on presences for each species (Batcheler 1971). Because wetlands in Yellowstone are significantly clustered from complete spatial randomness ($p < 0.001$), we evaluated whether species presence is more clustered than expected conditioned on the distribution of wetlands. To this end, we calculated the average distance to nearest neighbor with standard methodology but used an iterative random draw of selected sites as the null distribution (R core development team 2007). We evaluated spatial autocorrelation in species abundance using Moran's I (Moran 1950) and Geary's C (Geary 1954) with small distance correction in Crimestat 3.0 (Levine 2004). We investigated the strength and distance die-off of autocorrelation using a Moran's correlogram (1000 simulations, 200 bins, (Cliff and Haggett 1988)). Finally, we tested for significant differences between *P. maculata* and *B. boreas* spatial

distribution in Yellowstone using species abundance at each sample location and 1000 permutations (Syrjala 1996).

Grinellian niche

We evaluated the Grinellian niche of *P. maculata* and *B. boreas* by deriving variables of environmental condition that may define niche. We then developed modelling and evaluation methodology including: model selection, evaluation of model fit, validation of niche models for each species, and evaluation of difference in species' niche.

Variables of environmental condition

We modeled species niche based on environmental conditions: climate (temperature and rainfall), wetland characteristics, solar energy, dominant cover, site topography and water holding capacity, water chemistry, and disturbance (Table 2; for detailed variable description, source data, predicted effect, and ecological justification see Appendix 3). We derived independent variables from 30m spatial data: digital elevation model (DEM) derived from the spaceshuttle radar topography mission (SRTM), 2001 national landcover data (NLCD), and a spline-based climate model (Rehfeldt 2006, Rehfeldt et al. 2006) **(**Table 2; Appendix 3). In addition, we included wetland category and type from national wetland inventory (NWI) data for Yellowstone (Table 2, Appendix 3).

Model Parameterization

Using the variables in Table 2, we quantified species environmental requirements with RANDOM FORESTS. RANDOM FORESTS is a classifier based on a bootstrap of the data that grows many trees (Classification and Regression Trees – CARTs) where the prediction is based on votes across all tree nodes for the most popular (or likely) classification (Breiman 2001, Cutler et al. 2007). This powerful, non-parametric approach is able to incorporate complex interactions

among independent variables and outputs measures of variable importance (Déath and Fabricius 2000, Cutler et al. 2007). We executed RANDOM FORESTS in classification mode with 5000 trees, 34% out of bag (OOB) sample, *m* (number independent variables iterated at each node) set by the tuning parameter, and removal of multivariate redundant variable using qr matrix decomposition (threshold = 0.05) (Becker et al. 1988).

With rare species, presences may occur at a much lower rate than absences across surveyed sites. This results in an unbalanced sample, where one class (presence) is observed at a much lower rate than the other class (absences). Unbalanced samples are known to be a potential bias in RANDOM FORESTS models, especially for assessing classification accuracy (Svetnik et al. 2004). Therefore, we balanced the sample by iteratively subsampling absences at 2n the rate of presences (n). We subsampled the data, ran the model, and compared the covariance matrix (Morrison 2002) of the independent variables in the subsample to the independent variables of the whole dataset. We then repeated the process until the covariance matrices from the whole and subsampled datasets converged at p=0.05, thus representing the true distribution of independent variables. We then combined all trees into one model to calculate a final votes matrix, assess model fit, and calculate variable importance. This created a suite of trees that have balanced sample and yet still have a true bootstrap of the data unlike approaches that oversample the underrepresented class (Svetnik et al. 2004).

Model selection and fit

As an output of RANDOM FORESTS, variables are ranked in order of importance based on improvement of the model (IM, calculated by predicted OOB accuracy and accuracy across permuted variables at a node normalized by standard error). We ran an initial model with all variables, and then calculated a model improvement ratio (MIR). The MIR is a standardized

measure of model improvement given by [IM_n / IM_{max}]. We selected a MIR threshold (*n*, tested at 0.1 increments) and a corresponding model retaining all variables above the threshold, based on overall OOB error and smallest maximum class error. The overall OOB error is the median cross-classification rate for both classes (presence and absence).

We assessed model fit by overall model significance, overall OOB error and OOB error by class (presence vs. absence). To assess overall model significance, we created a null distribution by randomizing observed presences and calculating percent variation explained for each randomization (n=1000). We then calculated whether the percent variation explained for a given model was > 95th percentile of the null distribution (p-value < 0.05) (Murphy et al. in review). However, overall OOB error may be relatively low for one class but high for the other class. Therefore, we also assessed OOB error by class (presence vs. absence). We also evaluated the ability to separate presence-absence in multivariate space using multidimensional scaling (MDS) of the RANDOM FORESTS node proximity matrix.

Model validation

We validated selected models by fixing the parameters, excluding 10% of the data, refitting the votes matrix, and predicting to the withheld data (n=100 iterations). Based on the validation, we calculated overall percent of observations correctly classified (PCC), PCC by class, Kappa (Landis and Koch 1977) and the area under curve (AUC) using the Presence package in R (Presence/Absence package, R). To incorporate probability of detection < 1 (MacKenzie et al. 2002, MacKenzie et al. 2003), we reclassified presences as absences at 1-rate of detection estimated from monitoring surveys conducted in the same years (D. Patla, pub. comm.). We iterated this analysis 100 times for each species, rerunning the model each time to assess the effect of detection rate on stability of the validation statistics.

Grinellian niche separation

Based on the final environmental requirements model for each species, we assessed Grinellian niche dissimilarity by reciprocally predicting *B. boreas* presence using the model developed for *P. maculata* and vice versa. If these species have similar environmental requirements, a niche model developed for one species should be able to predict the niche of the other species. For each niche model, we exchanged the observed presences/absences with those for the other species and recalculated the fit and validation statistics.

Niche-based species distribution models

To predict dispersal limited niche, we added variables measuring dispersal limitation to the selected Grinellian niche model for each species. To account of the availability of a source population, we calculated distance to nearest occupied site scaled by the minimum distance to a surveyed site for each observation (Table 2, Appendix 3). We also included spatial coordinates as independent variables to account for regional trends. Finally, dispersal into suitable habitats may be limited environmental condition of the surrounding landscape. Therefore, we calculated the percent landscape for cover and disturbance variables at 200, 400, and 1000m radii around each site (PLAND, Fragstats (McGarigal and Marks 1995), Table 2). We followed model parameterization, selection, fit, and validation from the Grinellian niche method outline above. We compared model fit for Grinellian niche versus dispersal limited niche within a species by fit statistics, validation statistics, and spatial prediction. By treating the votes matrix as a probability density function, we predicted probability of occurrence for the landscape for both the Grinellian and dispersal limited niche distribution by passing a specialized function to the grid ASCII predict function in yaImpute (Crookston & Finley 2008) in R. This predicted species

distribution model based on each niche concept. We then compared the spatial predictions of Grinellian versus dispersal limited niche.

RESULTS

Field sampling

From 2004-2005, 378 sites were visited at least once during the amphibian breeding season (May-July). *Pseudacris maculata* was observed at 43.39% (164/378) of the selected sites; 6.88% (26/378) at low abundance, 8.20% (31/378) at medium abundance, and 28.04% (106/378) at high abundance (Fig 7). *Bufo boreas* was observed at 12.43% (42/378) of the selected sites; 5.82% (22/378) at low abundance, 3.17% (12/378) at medium abundance, and 3.44% (13/378) at high abundance (Fig. 7). Both species were present at 1.59% (6/378) of the sites.

Distribution

We found spatial distributions of *P. maculata* and *B. boreas* to be significantly different. Both *P. maculata* and *B. boreas* are spatially clustered based on a standard nearest neighbor ($p < 0.01$, Table 3). However, when nearest-neighbor is based on the null distribution of wetlands, only *P. maculata* distribution is significantly clustered ($p < 0.05$). There was significant spatial autocorrelation in abundance for both species with both Moran's *I* and Geary's C (Table 3), rejecting the null hypothesis of complete spatial randomness. *Pseudacris maculata* was more highly spatially autocorrelated ($I = 0.175$, $p < 0.001$) than *B. boreas* ($I = 0.054$, $p < 0.05$) across the whole study area (Table 3). Both species were significantly autocorrelated across all distance bins ($p < 0.05$, Appendix 4). However, *P. maculata* was more highly spatially autocorrelated across the range and degree of spatial autocorrelation declined more slowly than *B. boreas* (Appendix 4). Finally, spatial distribution of *P. maculata* and *B. boreas* presences are significantly different from each other (psi – 3.982, p-value 0.034).

Grinellian niche

Pseudacris maculata and *B. boreas* breeding niche can be separated based on environmental requirements. *Pseudacris maculata* breeding niche in Yellowstone is defined by temperature (degree days > 5° C, frost free period), wetland characteristics (type and category), moisture (mean annual precipitation, growing season precipitation), cover (shrub), and absence of thermal activity ($p < 0.05$, Table 4, MIR 0.7). We did not observe an imbalance in *P. maculata* presence (43.49%) versus absence observations, therefore we did not balance the sample. For model fit, OOB error was 21.1% for the whole model and 25.2% for the presence class (Table 4). When we validated the model, 98% of sites were correctly classified as presence/absence with a Kappa of 0.96 (Table 4). The two classes were visually separable on the MDS plot of the RANDOM FORESTS node proximity matrix (Appendix 4).

We only observed *B. boreas* at 12.43% of the sites. Initial models of *B. boreas* without balanced sample had > 60% error in the presence class but only 5% error in the absence class. Therefore, we followed the methods for an unbalanced sample. The *B. boreas* breeding niche in Yellowstone is defined by wetland characteristics (type), topography (elevation relief ratio at 27X 27 cell window size, topographic roughness at 27X27 cell window size), mean annual precipitation, surrounding wetland cover, and degree days > 5° C ($p < 0.05$, Table 4, MIR 0.6). For model fit, OOB error was 23.9% for the whole model, 31.8% for the presence class (Table 4). In model validation, we were able to correctly classify presence/absence for 95.8% of sites (presence 100%, absence 92.2%) with a Kappa of 0.915 (Table 4). Classes were only visually separable on a portion of the MDS plot (Appendix 4).

Grinellian niche models developed for each species performed poorly in the reciprocal prediction test, providing evidence of niche dissimilarity of these two species Yellowstone

(Table 4). When the *P. maculata* model was used to predict *B. boreas*, 4.2% were correctly classified on validation, and Kappa was -0.91 (Table 4). When the *B. boreas* model was used to predict *P. maculata*, 1.6% were correctly classified on validation, and Kappa was -0.968 (Table 4). Probability of detection < 1 had little effect on the results for either species (< 5% standard error), suggesting probability of detection was high enough that pseudo-absences had little effect (MacKenzie et. al 2003).

Niche-based species distribution models

Observed *P. maculata* and *B. boreas* distributions were best explained by dispersal limited niche theory. In our dispersal limited niche model for *P. maculata*, distance to nearest occupied site was the most important variable. After model selection, we also retained location (X) and cover (Shrub at 400m radius), in addition to all of the variables from the *P. maculata* Grinellian niche model (Table 4). For the niche-based distribution model, OOB error improved to 14.3% overall and 16.4% for the presence class (Table 4). Model validation also improved compared to the Grinellian model with 99.15 % correctly classified with a Kappa 0.983 (Table 4). As in the Grinellian model, the two classes were visually separable on the MDS plot (Appendix 4). The *P. maculata* distribution model based on the Grinellian niche lead to over-prediction of suitable habitat whereas the distribution model based on a dispersal limit niche was more restricted. Both distribution models identified Hayden Valley and wetlands north of Lewis Lake (Fig. 8) as high probability of species presence.

When we added variables measuring dispersal limitation to the *B. boreas* Grinellian niche model, distance to nearest occupied site was also the most important variable. We retained no other measures of dispersal limitation. With the inclusion of distance to nearest occupied site, OOB error improved to 20.9% overall and 22.9% for the presence class (Table 4). Model

validation also improved compared to the niche model with 97.35% correctly classified with a Kappa 0.947 (Table 4). The presence class had two clusters, one separable and one intermixed with absences (Appendix 4). In addition, the *B. boreas* niche distribution model still had higher OOB error in the presence class than the *P. maculata* Grinellian niche model (22.9% vs. 25.2%). The *B. boreas* distribution model based on the Grinellian niche led to slight over-prediction of suitable habitat whereas the distribution model based on a dispersal limited niche was more restricted. Both distribution models identified the areas around Old Faithful and east side of Lewis Lake (Figs. 7,8; Appendix 5) as areas with high probability of occupancy. In addition, we identified high probability of occupancy areas in the inaccessible southern arms of Yellowstone Lake (Fig. 8), where there are multiple historic observations (Koch and Peterson 1995).

DISCUSSION

We found *P. maculata* and *B. boreas* distributions were significantly different in Yellowstone with separable niches based on environmental requirements. However, species distribution models based on the Grinellian niche theory resulted in an over prediction of species presence. If our goal was to locate all habitats meeting basic environmental requirements, models based on the Grinellian niche theory would be preferred (Rehfeldt et al. 2006). However, the dispersal limited niche theory best explained the observed distributions. As part of our approach, we develop a cohesive framework that addresses current limitations of species distribution models (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Araujo and Guisan 2006, Austin 2007). For this reason, our models have high predictive power often lacking in amphibian distribution models (Green 2003, Denoel and Lehmann 2006).

Araujo and Guidan (2006) identify five challenges in species distribution modelling: clarification of the niche concept, improving study design, improving parameterization

(modelling) strategies, improving model selection and evaluation of variable contribution, and improving model validation. We successfully address these needs by i) testing competing niche theories and resulting predictions of distribution (Pulliam 2000, Austin 2007), ii) utilizing a stratified random sample (Guisan and Zimmermann 2000), iii) applying a non-parametric algorithmic approach (RANDOM FORESTS) with methodology for unbalanced samples and imperfect detection rates, iv) developing model selection criterion for RANDOM FORESTS that incorporate variable importance, and v) implementing a test of overall model significance while conducting both internal and external model validation. Using this modelling framework, we were able to build species distribution models based on ecological theory with high classification accuracy (>90%) for both *P. maculata* and *B. boreas*.

Distribution

Testing for a significant difference in the observed spatial distributions of two species is an important first step before proceeding to infer differentiation in species' distributions contingent on environmental requirements. Based on the location and abundance of observations for each species, we found the spatial distributions of *P. maculata* and *B. boreas* to be significantly different. Properties of the spatial distribution (e.g., autocorrelation, clustering) can also give insight into species' dispersal ability. Amphibian dispersal tends to be limited (Marsh et al. 1999, Manier and Arnold 2006)**,** therefore we expected occupied sites to be spatially autocorrelated for both species. While we detected spatial autocorrelation in occupied sites in both cases, stronger autocorrelation in *P. maculata* than *B. boreas* occupancy suggests *B. boreas* is capable of colonizing suitable habitat farther from occupied sites. This observation is supported by maximum recorded dispersal distances for each species (*P. maculata* 600 m (Spencer 1964), *B. boreas* 6000 m (Muths 2003)). In addition, *P. maculata* presence is

significantly clustered relative to the expected distribution based on wetland availability. This suggests suitable habitat remains unoccupied due to lack of a source population.

Grinellian niche

Although *P. maculata* and *B. boreas* are both pond-breeding amphibians with similar habitat requirements, we were able to differentiate their niches based on environmental conditions in Yellowstone (the "realized environment"). Our ability to effectively differentiate Grinellian niche was due in part to improvements in the modelling framework developed for this study. Using a random stratified sample, we were able to capture the variation in the temperature-moisture gradient. By balancing observations of species presence/absence, we were able to decrease error in predicting *B. boreas* presence. By incorporating uncertainty due to potential pseudo-absences, we demonstrated that our model results were insensitive to pseudo-absences. RANDOM FORESTS already contains a bootstrap of the data, so presence at a single site is unlikely to have disproportionate influence across replicate trees and therefore the final model. Finally, our Grinellian niche models have high (>92%) classification accuracy when validated against withheld data.

Pseudacris maculata

Pseudacris spp. are often through to be opportunistic with regard to breeding site selection (Stevens et al. 2007), laying eggs in any wetland with emergent vegetation for egg deposition (Koch and Peterson 1995). However, we found *P. maculata* in Yellowstone to be restricted based on environmental conditions with a definable Grinellian niche as seen both in the selected variables and PCC (98%). As demonstrated by the clear separation between presence and absence in the MDS plot (Appendix 4), *P. maculata* is not equally probable across the range of environmental variability as expected with an opportunistic species. There are two explanations

for this result: throughout its range *P. maculata* is not an opportunistic species as previously believed or the environmental conditions at the edge of the species' range represent only a marginal portion of the species' fundamental niche (Braunisch et al. 2008).

In the extreme environment of Yellowstone, two competing forces limit environmental suitability for *P. maculata*: wetland evaporation and length of the growing season (Corn 2003). In order to survive, *P. maculata* larvae must reach metamorphosis is a single season before temporary wetlands evaporate and subsequently larva must grow large enough to survive the harsh winter (Semlitsch et al. 1999, Beck and Congdon 2000). The effect of wetland evaporation, as measured by degree days > 5° C and wetland type, may be amplified by the long-term drought in Yellowstone (>8 year at time of data collection). In the most arid section of the park (Lamar valley, NE), we observed *P. maculata* at only two sites even though our surveys included multiple locations with historic breeding records (Koch and Peterson 1995).

Short growing seasons with cool temperatures, as measured by frost-free period, slow time to metamorphosis (Palo et al. 2003). We did not find *P. maculata* above 2654m in elevation, likely due to short frost-free period (average < 3 days). In addition, occupied sites above 2500m tended to be in areas with warmer temperatures and less cover then unoccupied sties at these elevations. Run-off from thermal pools elevates water temperature, which would be beneficial for *P. maculata* at sites with a short frost-free period. However, we observed a negative relationship between *P. maculata* presence and thermal influence. Thermal influence also changes water chemistry (e.g., conductivity) which may negatively impact *P. maculata* eggs, larvae, or vegetation required for egg deposition.

Occupancy of suitable habitats is further limited by competition and predation. Wetland category identified shallow, inundated wetlands with low current and vegetation required for *P.*

maculata breeding success (palustrine category). However, wetland category may also be an indirect measure of predation and competition. *Rana luteiventris* breeds in deeper, cooler wetlands and there may be competitive exclusion with *P. maculata.* In addition, many riverine (river) and lacustrine (lake-like) wetlands contain fish predators, which preclude *P. maculata* breeding (Pilliod and Peterson 2001).

Bufo boreas

We were able to define *B. boreas* Grinellian niche with high classification accuracy (PCC 95.5%). Although *B. boreas* is thought to have more specific breeding requirements than *P. maculata*, species' Grinellian niche was less definable. The MDS plot for *B. boreas* shows two clusters of presences intermixed with absences (Appendix 4). There are three alternative explanations for our result: *B. boreas* is a more opportunistic breeder than previously thought, the two breeding strategies (ponds vs. rivers) form a bimodal niche distribution, or the sample size is insufficient to clearly separate *B. boreas* presence/absence. Although our sample fairly represents the rarity of *B. boreas* in the study area by including both a random stratification and directed sampling, it may still lack the size needed for clear MDS clusters.

Environmental conditions related to site specific characteristics such as muddy, flat wetlands with slow moving water are highly important in defining *B. boreas* niche (Table 4). Although *B. boreas* breeds in both ponds and rivers, the utilized wetland types had similar muddy substrates that provide cover for larvae and breeding adults. We also found topographic measures related to broad-flat areas (elevation relief ratio and topographic roughness) were important. Elevation relief ratio is higher on slope angles, which are unlikely to contain these habitats. Topographic roughness is low in flatter areas such as thermal basins around Old Faithful which have a relatively high density of *B. boreas* breeding sites (Fig. 7).

Although not as important as in defining *P. maculata* niche, rainfall and temperature are driving variables for *B. boreas* niche. Moisture is essential for *B. boreas* hydration, as the species has little physiological control over water loss (Bartelt and Peterson 2005). However, mean annual precipitation is in the selected model as opposed to the additional inclusion of growing season precipitation as in the case of *P. maculata*. *Bufo boreas* river sites are not ephemeral and the wetlands associated with muddy flats are often fed by thermal springs which maintain some water in the wetland throughout the season. In addition, while degree days >5° is retained in the *B. boreas* model, it is not the most important variable as for *P. maculata*. *Bufo boreas* is highly terrestrial in the late summer and may have behavioral adaptations for avoiding water loss (Bartelt et al. 2004).

Niche-based species distribution models

Species distribution models based on a Grinellian niche alone resulted in over-prediction of species' fine-scale distribution (Fig. 8, Appendix 5). If dispersal is limited, as in both our focal organisms (Spencer 1964, Muths 2003, Manier and Arnold 2006), we would expect some habitats meeting species' environmental requirements to be unoccupied (Pulliam 2000). Therefore, observed *P. maculata* and *B. boreas* distributions are best explained by dispersal limited niche theory with distance to nearest occupied site the most important predictor. In addition, predicting probability of occurrences identifies the suitability of the habitat and certainty of the estimate giving a clearer understanding of distribution than presence-absence only predictions.

Pseudacris maculata is more spatially dependant than *B. boreas* as reflected in spatial autocorrelation of presences, significant clustering, and the including of additional spatial variables in the dispersal limited niche model (UTM_X and percent shrub cover, 400m window).

Location (UTM_X) indicated a regional effect on the east to west axis where sites in the central region of Yellowstone were more likely to be occupied. Drought conditions in the Lamar valley (northeastern section of the park) were particularly severe and may be drive this relationship. Finally, shrub cover at 400m radius was also important for *P. maculata* niche-based distribution emphasizing the importance of this cover type for site occupancy.

Species distribution model predictions based on the dispersal limited niche are more constrained than Grinellian niche species distribution model predictions (Fig. 8). The dispersal limited niche species distribution models more effectively reflect both regional variation and fine-scale occupancy. In addition, we were able to predict high probability of occupancy in unsampled areas supported by historic observations.

CONCLUSIONS

Species distribution models have many applications in ecology and conservation such as quantifying the environmental niche of a species, testing multiple ecological hypotheses, or providing guidance for directed sampling of rare or difficult to find species (Guisan and Thuiller 2005, Guisan et al. 2006). The prediction of probability of occurrence can identify hotspots (Garcia 2006) and will be useful in directing field surveys to sites likely to have breeding activity, especially for *B. boreas* (Appendix 5). In addition, the species distribution model can be projected forward in time under different scenarios of climate change (Rehfeldt et al. 2006) which may greatly impact amphibians (Araujo et al. 2006). Finally, the niche models will be used to inform future landscape genetics models assessing connectivity for both species (Manel et al. 2003, Storfer et al. 2007).

ACKNOWLEDGEMENTS

Research was permitted by Yellowstone National Park (Permit YELL-05452**;** C Smith and C Hendrix) and IACUC (ASAF #3378). We would like to thank M Murphy, M Flisher, D Murphy, D. Patla, C. Cockran, and D. Cockran for field assistance. A Gallant, P Bartelt, and R Klaver assisted in compiling spatial data. D. Patla, C. Goldberg, and D. Pilliod provided valuable advice and training for field sampling. Funding for MAM was provided by EPA-STAR fellowship (FP-916695**)**, NSF (grant DEB-0608458 to MAM and AS**),** James King fellowship, Theodore Roosevelt Memorial Fund, Society for Wetland Science (2004, 2005), WSU Zoology alumni scholarship, Canon National Parks, and Graduate Women in Science. AS was supported by NSF DEB-0548415.

LITERATURE CITED

Ackerly, D. D. 2003. Community assembly, niche conservatism and adaptive evolution in changing environments. International Journal of Plant Sciences 164:S164-S184.

Alonso, D., R. S. Etienne, and A. J. McKane. 2006. The merits of neutral theory. Trends in Ecology & Evolution 21:451-457.

Araujo, M. B., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography 33:1677-1688.

Araujo, M. B., W. Thuiller, and R. G. Pearson. 2006. Climate warming and the decline of amphibians and reptiles in Europe. Journal of Biogeography 33:1712-1728.

Arens, P., T. van der Sluis, W. P. van't Westende, B. Vosman, C. C. Vos, and M. J. M. Smulders. 2007. Genetic population differentiation and connectivity among fragmented Moor frog (*Rana arvalis*) populations in The Netherlands. Landscape Ecology 22:1489-1500.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling 200:1-19.

Babbit, K. J., M. J. Baber, and L. A. Brandt. 2006. The effect of woodland proximity and wetland characteristics on larval anuran assemblages in an agricultural landscape. Canadian Journal of Zoology 84:510-519.

Bartelt, P., and C. R. Peterson. 2005. Physical modeling operative temperature and evaporative water loss in amphibians. Journal of Thermal Biology 30:93-102.

Bartelt, P., C. R. Peterson, and R. W. Klaver. 2004. Sexual differences in the movements and habitats selected by Western toads in southeastern Idaho. Herpetologica 60:455-467.

Batcheler, C. L. 1971. Estimation of Density from a Sample of Joint Point and Nearest-Neighbor Distances. Ecology 52:703-709.

Beck, C. W., and J. D. Congdon. 2000. Effects of age and size at metamorphosis in *Hyla squirella*. Oecologia 112:187-192.

Becker, R., J. Chamers, and A. Wilks. 1988. The new S language: a programming environment for data analysis and graphics. Murry Hill, New Jersey.

Bosch, J., and I. Martinez-Solano. 2003. Factors influencing occupancy of breeding ponds in a montane amphibian assemblage. Journal of Herpetology 37:410-413.

Braunisch, V., K. Bollmann, R. Graf, and A. Hirzel. 2008. Living on the edge - Modelling habitat suitability for species at the edge of their fundamental niche. Ecological Modelling 214:153-167.

Breiman, L. 2001. Random Forests. Machine Learning 45:5-32.

Carey, C. 1993. Hypothesis concerning the causes of the disappearance of boreal toads from the mountains of Colorado. Conservation Biology 7:355-362.

Cliff, A. D., and P. Haggett. 1988. Atlas of disease distributions. Blackwell Reference, Oxford.

Corn, P. S. 2003. Amphibian breeding and climate change: importance of snow in the mountains. Conservation Biology 17:622-625.

Cutler, D. R., T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

Déath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178-3192.

Denoel, M., and A. Lehmann. 2006. Multi-scale effect of landscape processes and habitat quality on newt abundance: Implications for conservation. Biological Conservation 130:495-504.

Evans, I. S. 1972. General geomorphometry, detivatives of altitude, and descriptive statistics. *in* R. J. Chorley, editor. Spatial Analysis in geomorphology. Harper & Row, New York.

Fu, P., and P. M. Rich. 1999. Design and implementation of the Solar Analyst: an ArcView extension for modeling solar radiation at landscape scales. *in* 19th Annual ESRI User Conference. ESRI, San Diego, CA.

Funk, W. C., M. S. Blouin, P. S. Corn, B. A. Maxell, D. S. Pilliod, S. Amish, and F. W. Allendorf. 2005. Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. Molecular Ecology 14:483-496.

Garcia, A. 2006. Using ecological niche modeling to identify diversity hotspots for the herpetofauna of Pacific lowlands and adjacent interior valleys of Mexico. Biological Conservation 130:25-46.

Gardner, T. A., J. Barlow, and C. A. Peres. 2007. Paradox, presumption and pitfalls in conservation biology: The importance of habitat change for amphibians and reptiles. Biological Conservation 138:166-179.

Geary, R. C. 1954. The contiguity ratio and statistical mapping. The Incorporated Statistician 5:115-145.

Giordano, A. R., B. J. Ridenhour, and A. Storfer. 2007. The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactylum*). Molecular Ecology 16:1625-1637.

Goates, M. C., K. A. Hatch, and D. L. Eggett. 2007. The need to ground truth 30.5 m buffers: A case study of the boreal toad (*Bufo boreas*). Biological Conservation 138:474-483.

Green, D. M. 2003. The ecology of extinction: population fluctuation and decline in amphibians. Biological Conservation 111:331-343.

Guisan, A., O. Broennimann, R. Engler, M. Vust, N. G. Yoccoz, A. Lehmann, and N. E. Zimmermann. 2006. Using Niche-Based Models to Improve the Sampling of Rare Species. Conservation Biology 20:501-511.

Guisan, A., and W. Thuiller. 2005. Predicting species distributions: offering more than simple habitat models. Ecology Letters 8:993-1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modeling 135:147-186.

Hossack, B. R., and P. S. Corn. 2007. Responses of pond-breeding amphibians to wildfire: short-term patterns in occupancy and colonization. Ecological Applications 17:1403-1410.

Koch, E. D., and C. R. Peterson. 1995. Amphibians and Reptiles of Yellowstone and Grand Teton National Parks. University of Utah Press, Salt Lake City.

Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159-174.

Levine, N. 2004. CrimeStat III: A spatial statistics program for the analysis of crime incident locations. National Institute of Justice, Washington DC, Houston, TX.

Lougheed, S. C., C. Gascon, D. A. Jones, J. P. Bogard, and P. T. Boag. 1999. Ridges and rivers: a test of competing hypotheses of Amazonian diversification using a dart-poison frog (*Epipedobates femoralis*). Proceedings of the Royal Society B 266:1829-1835.

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84:2200-2207.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248-2255.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189-197.

Manier, M. K., and S. J. Arnold. 2006. Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. Proceedings of the Royal Society B 273:3001-3009.

Marsh, D. M., E. H. Fegraus, and S. Harrison. 1999. Effect of breeding pond isolation on the spatial and temporal dynamics of pond use by the tungera frog, *Phylsalaemus pustulosus*. Journal of Animal Ecology 86:804-814.

McGarigal, K., and B. J. Marks. 1995. FRAGSTATS. Spatial analysis program for quantifying landscape structure. USDA Forest Service General Technical Report PNW-GTR-351.

Moore, I., P. Gessler, G. Nielsen, and G. Petersen, editors. 1993. Terrain attributes and estimation methods and scale effects. Wiley, London.

Moran, C. 1950. Notes on continuous stochastic phenomena. Biometrika 37:17-23.

Moriarty Lemon, E. C., A. R. Lemmon, and D. C. Cannatella. 2007. Geological and climatic forces driving speciation in the continentally distributed trilling chorus frogs (*Pseudacris*). Evolution 61:2086-2103.

Morrison, D. 2002. Multivariate Statistical Methods, 4th edition. McGraw-Hill, New York.

Munger, J. C., M. Gerber, K. Mardrid, M.-A. Carroll, W. Patersen, and L. Heberger. 1998. U.S. National Wetland Inventory classifications as predictors of the occurrence of Columbia

spotted frogs (*Rana luteiventris*) and Pacific treefrogs (*Hyla regilla*). Conservation Biology 12:320-330.

Murphy, M. A., J. S. Evans, and A. Storfer. in review. Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. Ecology.

Muths, E. 2003. Home range and movements of boreal toads in undisturbed habitat. Copeia 1:160-165.

Muths, E., P. S. Corn, A. P. Pessier, and D. E. Green. 2003. Evidence for disease-related amphibian decline in Colorado. Biological Conservation 110:357-365.

Palo, J. U., R. B. O'Hara, A. T. Laugen, A. Laurila, C. R. Primmer, and J. Merila. 2003. Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. Molecular Ecology 12:1963–1978.

Pawar, S., M. S. Koo, C. Kelley, M. F. Ahmed, S. Chaudhuri, and S. Sarkar. 2007. Conservation assessment and prioritization of areas in Northeast India: Priorities for amphibians and reptiles. Biological Conservation 136:346-361.

Pechmann, J. H. K., D. E. Scott, J. W. Gibbons, and R. D. Semlitsch. 1989. Influence of wetland hydroperiod on diversity and abundance of metamorphosing juvenile amphibians. Wetlands Ecology and Management 1:3-11.

Pilliod, D. S., and C. R. Peterson. 2001. Local and landscape effects of introduced trout on amphibians in historically fishless watersheds. Ecosystems 4:322-333.

Pilliod, D. S., C. R. Peterson, and P. I. Ritson. 2002. Seasonal migration of Columbia spotted frogs (*Rana luteiventris*) among complementary resources in a high mountain basin. Canadian Journal of Zoology 80:1849–1862.

Pulliam, H. R. 2000. On the relationships between niche and distribution. Ecology Letters 3:349-361.

R Development Core Team. 2007. R: A language and environment for statistical computing, version 2.7.0. R Foundation for Statistical Computing, Vienna, Austria.

Rehfeldt, G. E. 2006. A spline model of climate for the Western United States. RMRS-GTR-165, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Rehfeldt, G. E., N. L. Crookston, M. V. Warwell, and J. S. Evans. 2006. Empirical analysis of plant-climate relationships for western United States. International Journal of Plant Sciences 167:1123-1150.

Riley, S. J., S. D. DeGloria, and R. Elliot. 1999. A terrain ruggedness index that quantifies topographic heterogeniety. Intermountain Journal of Sciences 5:23-27.

Rose, C. S. 2005. Integrating ecology and developmental biology to explain the timing of frog metamorphosis. Trends in Ecology & Evolution 20:129-135.

Semlitsch, R. D. 2002. Critical elements for biologically based recovery plans of aquatic-breeding amphibians. Conservation Biology 16:619-629.

Semlitsch, R. D., J. Pickle, M. J. Parris, and R. D. Sage. 1999. Jumping performance and short-term repeatability of newly metamorphosed hybrid and parental leopard frogs (*Rana sphenocephala* and *Rana blairi*). Canadian Journal of Zoology 77:748-754.

Sinsch, U. 1988. Seasonal changes in the migratory behavior of the toad *Bufo bufo*: direction and magnitude of movements. Oecologia 76:391-398.

Smith, A. M., and D. M. Green. 2005. Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations? Ecography 28:110-128.

Spencer, A. W. 1964. The relationship of dispersal and migration to gene flow in the boreal chorus frog. PhD. Colorado State University.

Stage, A. 1976. An expression of the effects of aspect, slope, and habitat type on tree growth. Forest Sciences 22:457-469.

Stevens, C. E., C. A. Paszkowski, and A. L. Foote. 2007. Beaver (Castor canadensis) as a surrogate species for conserving anuran amphibians on boreal streams in Alberta, Canada. Biological Conservation 134:1-13.

Storfer, A. 2003. Amphibian declines: future directions. Diversity and Distributions 9.

Storfer, A., M. A. Murphy, J. S. Evans, C. Goldberg, S. Spear, S. Robinson, R. Dezzani, E. Demmelle, L. Vierling, and L. P. Waits. 2007. Putting the 'landscape' in landscape genetics. Heredity 98:128-142.

Stuart, S. N., J. S. Chanson, N. A. Cox, B. E. Young, A. S. L. Rodrigues, D. L. Fischman, and R. W. Waller. 2004. Status and trends of amphibian declines and extinctions worldwide. Science 306:1783-1786.

Svetnik, V., A. Liaw, C. Tong, and T. Wang. 2004. Applications of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. Pages 334-343 *in* F. Roli, J. Kittler, and Windeatt, editors. Lecture Notes in Computer Science. Springer, Berlin.

Syrjala, S. E. 1996. A statistical test for a difference between the spatial distributions of two populations. Ecology 77:75-80.

Tobler, W. R. 1979. Smooth pycnophalactic interpolation for geographical regions. Journal of the American Statistical Association 74:519-530.

Werner, E. E., D. K. Skelly, R. A. Relyea, and K. L. Yurewicz. 2007a. Amphibian species richness across environmental gradients. Oikos 116:1697-1712.

Werner, E. E., K. L. Yurewicz, D. K. Skelly, and R. A. Relyea. 2007b. Turnover in an amphibian metacommunity: the role of local and regional factors. Oikos 116:1713-1725.

Zeverbergen, L. W., and C. R. Thorne. 1987. Quantitative analysis of land surface topography. Earth Surface Processes and Landforms 12:47-56.

Table 2

Independent variables. Table 2 includes variables measuring environmental or dispersal limitation conditions. Expanded version of Table 2 with data source, variable calculation, expected relationship, and ecological justification is found in Appendix 3.

Condition	Variables	Code
Climate	Degree days > 5° C	dd5
	Frost free period	ffp
	Growing season precipitation	gsp
	Mean annual precipitation	map
Wetland characteristics	Wetland category	category
	Wetland Type	type
Solar energy	Heat Load Index	hli
	Hot-dry slopes	ssina
	Solar insolation	inso
	Canopy Percent[1]	canopy
Cover	Water[1]	water
	Forest[1]	forest
	Wetlands[1]	wetland
	Meadow[1]	meadow
	Barren[1]	barren
Water holding capacity	Compound topographic index	cti
	Curvature	crv
Topography	Topographic roughness[2]	rough
	Elevation relief ratio[2]	err
	Relative slope position	rsp
	Hierarchical slope position	hsp
Disturbance	1988 Fire perimeter[1]	burn
	Impervious surfaces[1]	imperv
Water Chemistry	Thermal[1]	therm
Isolation	Distance to nearest occupied site	distance
Regional effects	Location	X, Y
Dispersal cover	Water[3]	water
	Forest[3]	forest
	Wetlands[3]	wetland
	Meadow[3]	meadow
	Barren[3]	barren

Some variables were calculated at multiple window sizes [1] Percent of landscape (PLAND) in a 100m radius around the sample location [2]Calcualted at 3X3, 15X15, and 27X27 cell window sizes. [3]PLAND in 200, 400, and 1000m radii around the sample location.

Table 3

Spatial clustering and autocorrelation of P. maculata and B. boreas. NN is the standard nearest neighbor cluster statistic. Relative NN is the modified NN statistic based on the distribution of wetlands. Moran's I and Geary's C are measures of spatial autocorrelation.

Species	NN	Relative NN	Moran's I	Geary's C
P. maculata	0.21**	0.18*	0.175***	0.894***
B. boreas	0.25**	0.11	0.0535*	0.946*

*Significant 0.05 **Significant < 0.01 *** Significant < 0.001

Table 4

Model Results for Grinellian and dispersal limited niche models. Model is the species for which the model was fit (PSMA - *P. maculata*; BUBO - *B. boreas*) and niche theory used (G – Grinellian, DL – dispersal limited). Species – species predicted with a given model. Model fit – out of bag error (OOB) give as: overall model (absence, presence). Validation statistics - percent correctly classified (PCC: overall model (absence, presence)), Kappa, and area under curve (AUC). Selected model lists variables in order of importance. Abbreviations and explanations of variables are in Table 2.

		Model Fit	**Validation**			**Selected Model**
Model	Species	OOB	PCC	Kappa	AUC	Variables ranked by importance
PSMA-G	PSMA	21.1 (18.2, 25.2)	98.0 (97.8, 98.3)	0.960	0.980	dd5, wetland type, ffp, map, shrub(100),
	BUBO	Na	1.6 (1.2, 1.9)	-0.968	-0.984	wetland category, gsp, therm(200)
BUBO-G	BUBO	23.9 (11.5,31.8)	95.8 (92.2, 100)	0.915	0.957	wetland type, err27, map, wetland(100),
	PSMA	Na	4.2 (0, 4.2)	-0.915	-0.96	dd5, rough27
PSMA-DL	PSMA	14.3 (12.7,16.4)	99.2 (99.5, 98.8)	0.995	0.995	distance,dd5,wetland type, map, shrub(400), X, gsp, ffp, wetland category, therm(200)
BUBO-DL	BUBO	20.9 (11.1,22.9)	97.4 (95.5, 99.4)	0.947	0.964	distance, wetland type, map, dd5, err27, rough27, wetland(100)

FIGURE LEGENDS

Figure 7. Study area and observations. Figure 7 is a map of the study area with points represented surveyed locations overlaid on a shaded relief. Colors are as follows: black – no breeding presence for focal species, gray – *P. maculata* breeding presence, and white – *B. boreas* breeding presence. Dark areas represent major water bodies as a reference. Focal areas referred to in the text are labeled as follows: A) Yellowstone Lake, B) Hayden Valley, C), Lewis Lake, and D) Old Faithful geyser basin.

Figure 8. Species distribution models. Figure 8 displays the predicted probability of using spatial distribution models developed from each niche theory. Probability of occurrence scales from 0 (white) – 1 (black) predicted to the NWI data. Predicted distributions are as follows: A) *P. maculata* species distribution model based on the Grinellian niche, B) *P. maculata* species distribution model based on the dispersal-limited niche, C) *B. boreas* species distribution model based on the Grinellian niche, and D) *B. boreas* species distribution model based on the dispersal limited niche. Color versions of the figures and figures of focal areas are available in Appendix 5.

Figure 7

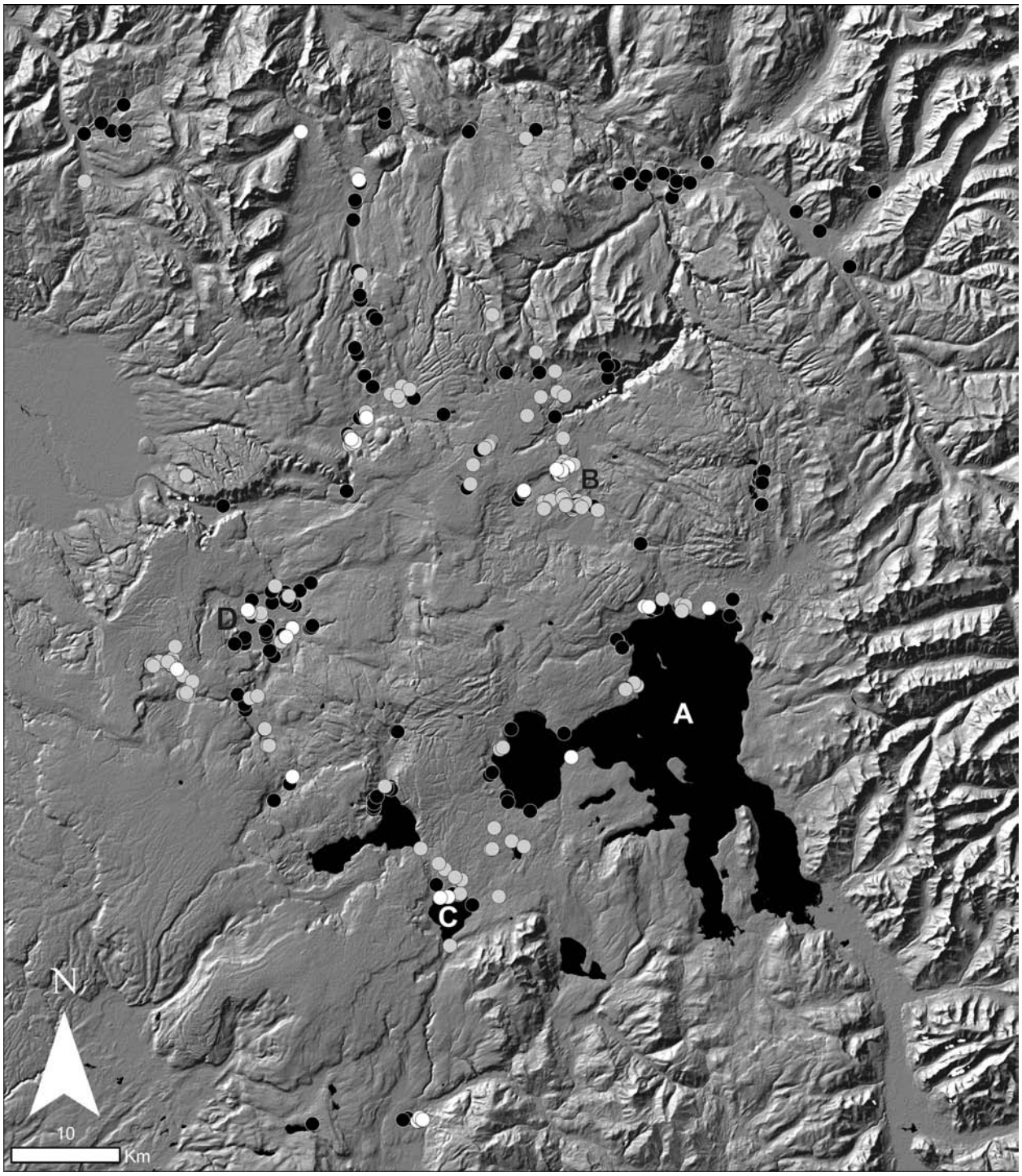
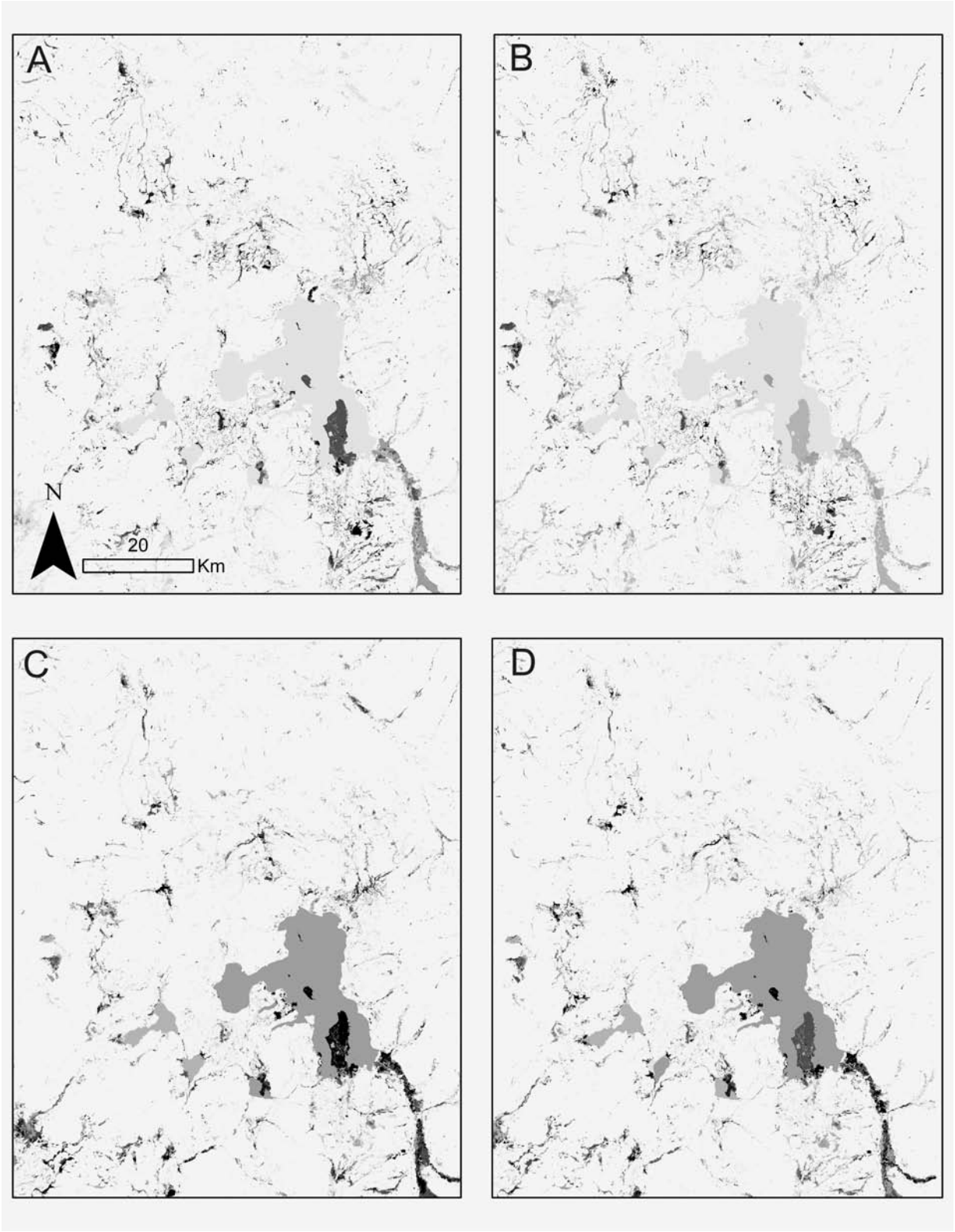


Figure 8



Appendix 3. Independent variables.

Appendix 3 is an expanded version of Table 2 from the text. Table headings are as follows. Condition is the environmental or dispersal condition that may define niche. Variables of these conditions and respective codes are listed in the next two columns. Source is the data source, either spatial data or model from which the data were derived. Pred is the predicted relationship between the variable of condition and species presence for each species (*P. maculata* - PSMA, *B. boreas* – BUBO). Plain text is a neutral relationship, **bold text** is a negative relationship, and *italicized text* is a positive relationship. Variable explanation is a short explanation of what the variable is measuring. Calculation gives either an explanation of how the variable was derived or a reference for the variable. Ecological justification is a justification for using the variable as a measure to define *P. maculata* and/or *B. boreas* niche including selected references.

Condition	**Variable**	**Code**	**Source**	**Pred**	**Variable Explanation**	**Calculation**	**Ecological Justification**
Climate	Degree days > 5° C	dd5	Spline[a]	BUBO **PSMA**	Sum of degrees for all days > 5° C	(Rehfeldt 2006)	High temperatures increase desiccation risk (Bartelt and Peterson 2005) and related to hydroperiod, important for breeding (Werner et al. 2007a)
	Frost free period	ffp	Spline	BUBO *PSMA*	Julian date of last freeze to date of first freeze	(Rehfeldt 2006)	Length of frost free period controls the length of the breeding season, which affects time to metamorphosis (Palo et al. 2003) and may correspond to areas more recently colonized (Moriarty Lemon et al. 2007).
	Growing season precipitation	gsp	Spline	*BUBO* *PSMA*	Average growing season precipitation	(Rehfeldt 2006)	Affects hydroperiod (Sinsch 1988, Werner et al. 2007a) and moisture.
	Mean annual precip	map	Spline	*BUBO* *PSMA*	Average annual precipitation	(Rehfeldt 2006)	Year-round precipitation, including snow pack, may influencing breeding timing and hydroperiod (Corn 2003, Werner et al. 2007a)
	Hot-dry slopes	ssina	SRTM[b]	**BUBO** **PSMA**	Indicates temperature-moisture	(Stage 1976)	High temperatures increase desiccation risk (Bartelt and Peterson 2005).
	Solar insolation	inso	SRTM	*BUBO* *PSMA*	Measure of solar energy at wetland Breeding: April – July	(Fu and Rich 1999)	Year – measure of site productivity. Breeding season solar insolation relates energy needed for tadpole rearing (Rose 2005)
	Canopy Percent[1]	canopy	NLCD[c]	**BUBO** PSMA	Percent cells forest cover in 3X3 window around wetland	PLAND (McGarigal and Marks 1995)	*B. boreas* prefers more open habitats (Bartelt et al. 2004)
Abiotic wetland characteristics	Wetland category	category	NWI	BUBO **PSMA**	Category of wetland (river, lake, pond)	NWI	*P. maculata* sensitive to fish predation, while *B. boreas* is toxic to many predators
	Wetland Type	type	NWI	BUBO **PSMA**	Permanence and substrate type	NWI	Breeds in shallow, inundated wetlands (Koch and Peterson 1995) which may be created by bison or beaver (Stevens et al. 2007).

Condition	Variable	Code	Source	Pred	Variable Explanation	Calculation	Ecological Justification
Cover	Water[1]	water	NLCD	**BUBO** *PSMA*	Open water	PLAND (McGarigal and Marks 1995)	May provide stopping locations during long-distance dispersal (Pilliod et al. 2002)
	Forest[1]	forest	NLCD	**BUBO** *PSMA*	Classes Deciduous, Evergreen, Mix	PLAND (McGarigal and Marks 1995)	Woodland proximity (Babbit *et al.* 2006), surrounding habitat important (Denoel and Lehmann 2006)
	Wetlands[1]	wetland	NLCD	BUBO *PSMA*	Classes Emergent & Herbaceous	PLAND (McGarigal and Marks 1995)	Number of sites in close proximity may be important for sustaining populations (Stevens et al. 2007).
	Meadow[1]	meadow	NLCD	*BUBO* PSMA	Classes: 71, 81, 82	PLAND (McGarigal and Marks 1995)	Provide shrub cover (Bartelt et al. 2004)
	Barren[1]	barren	NLCD	**BUBO PSMA**	Classes: 31, 12	PLAND (McGarigal and Marks 1995)	Little cover, exposed (Bartelt et al. 2004)
Holding capacity and topography	Compound topographic index	cti	SRTM	*BUBO PSMA*	Measure of wetness: flow accumulation by catchment size	(Moore et al. 1993)	Wetness may enhance dispersal (Bartelt and Peterson 2005, Goates et al. 2007) and influence breeding site hydroperiod (Pechmann et al. 1989).
	Curvature	crv	SRTM	*BUBO PSMA*	Curvature in elevation raster	(Zeverbergen and Thorne 1987)	High curvature may retain moisture while low curvature may be exposed.
	Topographic roughness[2]	rough	SRTM	BUBO **PSMA**	Variance in elevation for a given window size	(Riley et al. 1999)	Topographically complex areas may make dispersal energetically expensive
	Elevation relief ratio[2]	err	SRTM	BUBO **PSMA**	Index of elevational complexity	(Evans 1972)	Topographically complex areas may make dispersal energetically expensive
	Relative slope position	rsp	SRTM	**BUBO PSMA**	Relative position between valley floor and ridge top	(Murphy et al. in review)	Sites of similar slope position may be more connected
	Hierarchical slope position	hsp	SRTM	**BUBO PSMA**	Scale decomposition of slope position	(Murphy et al. in review)	Exposed areas may impede dispersal (Bartelt et al. 2004).
Disturbance	1988 Fire perimeter[1]	burn	USGS[e]	*BUBO* PSMA	Fire parameter from 1988	PLAND (McGarigal and Marks 1995)	Increased movement post fire (Hossack and Corn 2007)

	Impervious surfaces[1]	imperv	NLCD	**BUBO PSMA**	0-100% impervious surface	PLAND (McGarigal and Marks 1995)	Roads & development may limit dispersal ability (Arens et al. 2007)
Water Chemistry	Thermal[1]	therm	USGS	*BUBO* **PSMA**	Percent thermal cells	PLAND (McGarigal and Marks 1995)	Direct thermal is unsuitable habitat, but breed in areas with thermal influence. (Koch and Peterson 1995)
Regional Effects	UTM_X UTM_Y coordinates	XY	GPS[f]	BUBO *PSMA*	Location of site – gives similarity of location to other sites	None	Sites with more nearby occupied sites may be more likely to be occupied (Tobler 1979). Regional effect (Manier and Arnold 2006)
Isolation	Distance	distance	SRTM	BUBO *PSMA*	Distance to nearest occupied site	Topographically correct distance to nearest neighbor	Beyond the maximum dispersal distance, habitat may be unavailable (Spencer 1964). Local effect
Dispersal cover	Water[3]	water	NLCD	**BUBO** *PSMA*	Open water	PLAND (McGarigal and Marks 1995)	Water may provide stopping points during dispersal (Pilliod et al. 2002)
Condition	**Variable**	**Code**	**Source**	**Pred**	**Variable Explanation**	**Calculation**	**Ecological Justification**
	Forest[3]	forest	NLCD	**BUBO** *PSMA*	Classes Deciduous, Evergreen, Mix	PLAND (McGarigal and Marks 1995)	Woodland proximity (Babbit *et al.* 2006), surrounding habitat important (Denoel and Lehmann 2006)
	Wetlands[3]	wetland	NLCD	BUBO *PSMA*	Classes Emergent & Herbaceous	PLAND (McGarigal and Marks 1995)	Proximity to other potentially occupied sites, number of sites may be important (Stevens et al. 2007).
	Meadow[3]	meadow	NLCD	*BUBO* PSMA	Classes: 71, 81, 82	PLAND (McGarigal and Marks 1995)	Provide shrub cover (Bartelt et al. 2004)
	Barren[3]	barren	NLCD	**BUBO PSMA**	Classes: 31, 12	PLAND (McGarigal and Marks 1995)	Little cover, exposed (Bartelt et al. 2004)

Some variables were calculated at multiple window sizes [1] Percent of landscape (PLAND) in a 100m radius around the sample location [2]Calcualted at 3X3, 15X15, and 27X27 cell window sizes. [3] Percent of landscape (PLAND) in 200, 400, and 1000 m radii around the sample location. [a]Spline climate model (Rehfeldt 2006)[b]SRTM – Spaceshuttle radar telemetry mission data used to derive

a digital elevation model (DEM). [c]NLCD National Landcover dataset, 2001 [d] NWI – National Wetland Inventory [e]USGS – United States Geological Survey data for Yellowstone National Park. [f]GPS - Geographic Positioning System point location taken in the field.

Appendix 4 Additional figures for Moran's correlogram and MDS plots

Appendix 4 contains additional figures for a) Moran's correlogram and b) multidimensional scaling (MDS) plots.

Figure A4.1 Moran's correlogram. Figure A4.1 is a correlogram plotting Moran's *I* (Y) by distance bin (X, ~583 m/distance bin). *P maculata* values are represented by dashed lines while *B. boreas* values are represented by solid lines. Black lines are the observed Moran's *I* by distance bin. Gray lines are the upper bound for the 95% confidence interval simulation envelope.

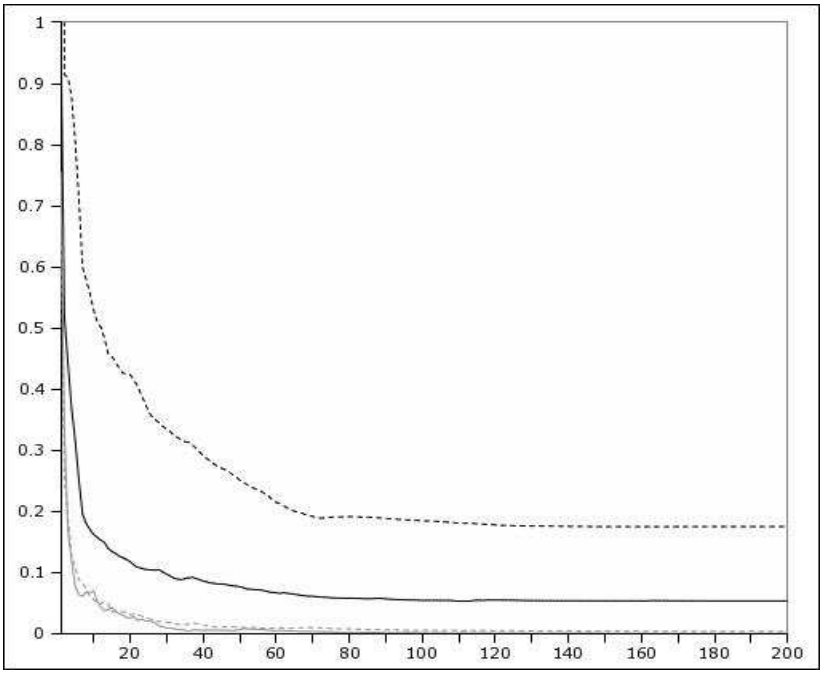
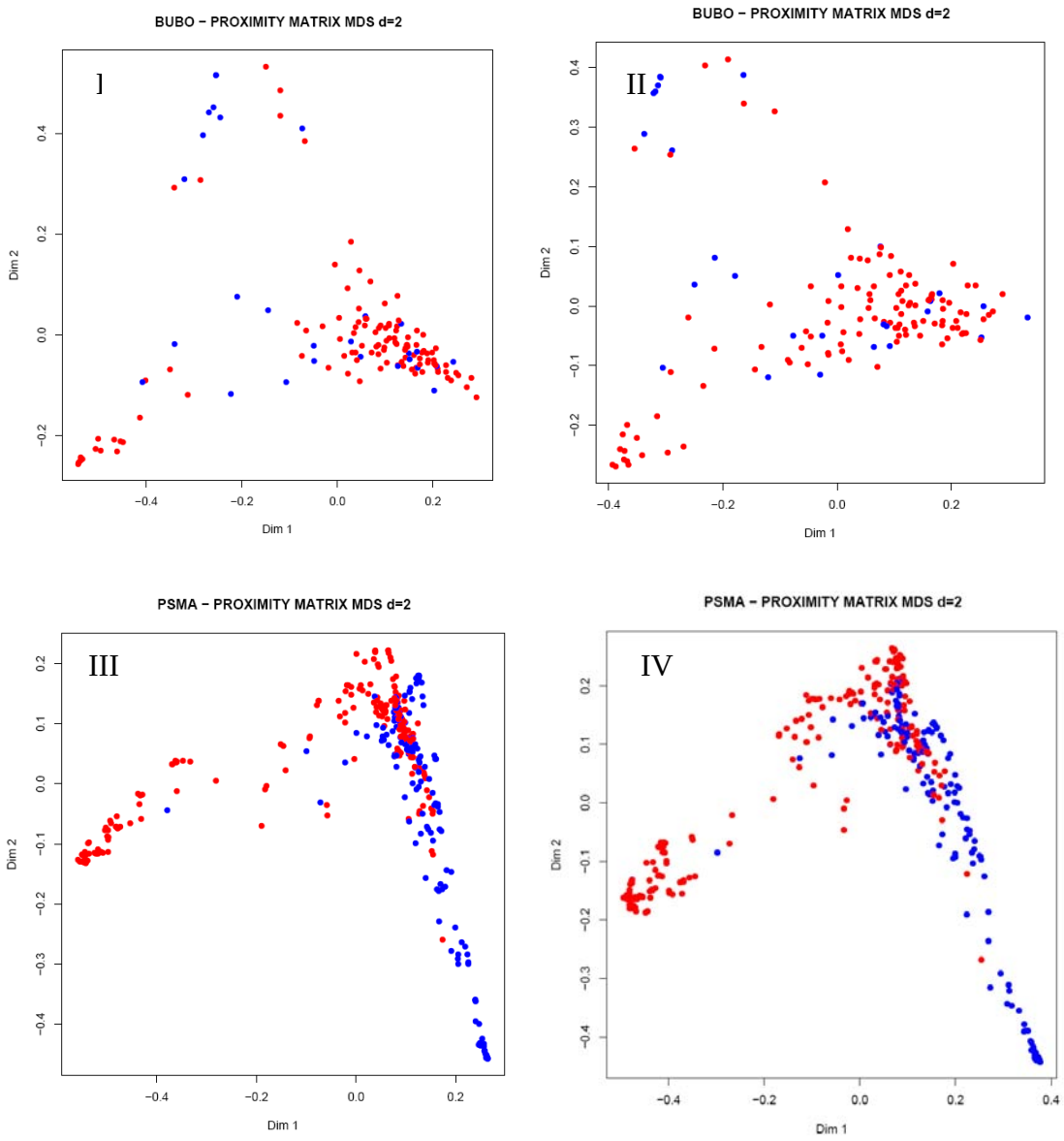


Figure A4.2. Figure A4.2 displays the multidimensional scaling (MDS) plots for each niche model. The MDS plots show separation between presences and absences by the nodes matrix. Panels are as follows: I) *B. boreas* MDS plot for the Grinellian niche mode, II) *B. boreas* MDS plot for the dispersal limited niche mode, III) *P. maculata* MDS plot for the Grinellian niche model, and IV) *P. maculata* MDS plot for the dispersal limited niche model. For all plots, presence is given in blue while absence is given in read.



Appendix 5. Color figures and focal distribution maps

Figure A5.1. Species distribution models (Color version of Fig. 8). Figure 8 displays the predicted probability of occurrence to the landscape using spatial distribution models developed from each niche theory. Probability of occurrence scales from 0 (white) – 1 (black) limited to the NWI wetlands. Predicted distributions are as follows: A) *P. maculata* species distribution model based on the Grinellian niche, B) *P. maculata* species distribution model based on the dispersal-limited niche, C) *B. boreas* species distribution model based on the Grinellian niche, and D) *B. boreas* species distribution model based on the dispersal limited niche.

Figure A5.2. Species distribution models by focal regions (see Fig. 7 for overall context). A) *Pseudacris maculata* species distribution models for Yellowstone Lake area. AI – *P. maculata* species distribution model based on Grinellian niche. The areas south of the lake are likely over predicted. AII – *P. maculata* species distribution model based on dispersal limited niche. Small wetlands in the southern portion of this focal area now have higher probability of presence, likely due open shrub habitat. The prediction is also more restricted, with fewer high probability sites on the south end of Yellowstone Lake. However, in the Grinellian niche model, an island in the northern portion of Yellowstone is predicted as high probability of breeding presence. Probability of presence drops with the dispersal limited niche model. This site is occupied (D. Patla, pers. com.), suggesting that there is a trade-off between over prediction of distribution and potential commission errors. Although this site is occupied, due to it's isolation the biological reality may be that it has a low probability of occupancy. B) Comparison of *P. maculata* (BI) and *B. boreas* (BII) species distribution

models (dispersal limited) for the Hayden Valley/Alum creek focal area. Although species presence in spatially intermixed, we are able to predict *P. macualta* versus *B. boreas* presence. In this area, *B. boreas* tends to breed in the river channel while *P. maculata* breeds in inundated wetlands. However, the two *B. boreas* non-river breeding sites are correctly classified as high probability for *B. boreas* and low probability for *P. maculata*. C) Species distribution model (dispersal limited) for *B. boreas* around Lewis Lake. Observations of *B. boreas* are marked with a cross symbol (all non-breeding), while unoccupied sites are denoted with a white dot. We did not observe any breeding site occupancy in this area and non-breeding observations were not included in the model. However, adjacent to the site where we observed a subadult (western location) was identified as high probability of breeding presence. All of the unoccupied sites had low probability of breeding presence. D) Species distribution model (distance limited) for *P. maculata* around Old Faithful. Most observations were well predicted (presence – cross symbol, absence white dot), even though presence-absences were in close proximity (nearly adjacent wetlands). There are a few commission errors on the E side of this focal area (wetlands with presence predicted to have low probability of breeding). This could be due to thermal influence in the area, negatively associated with *P. maculata* breeding. Although these wetlands are immediately adjacent to thermal areas, thermal run-off drains away from occupied sites into the Firehole River. The ability to include drainage patterns would likely improve model accuracy.

Figure A5.1

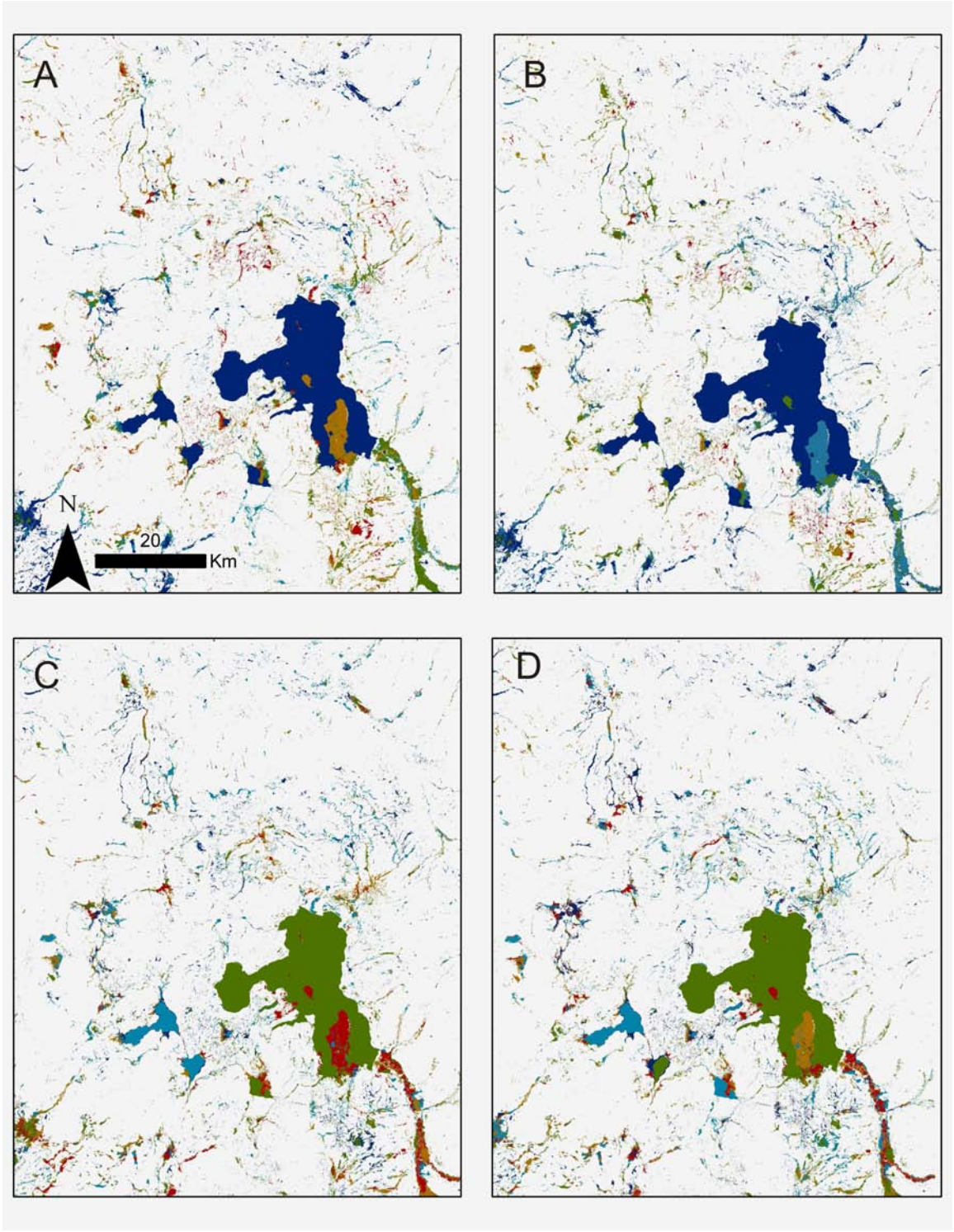
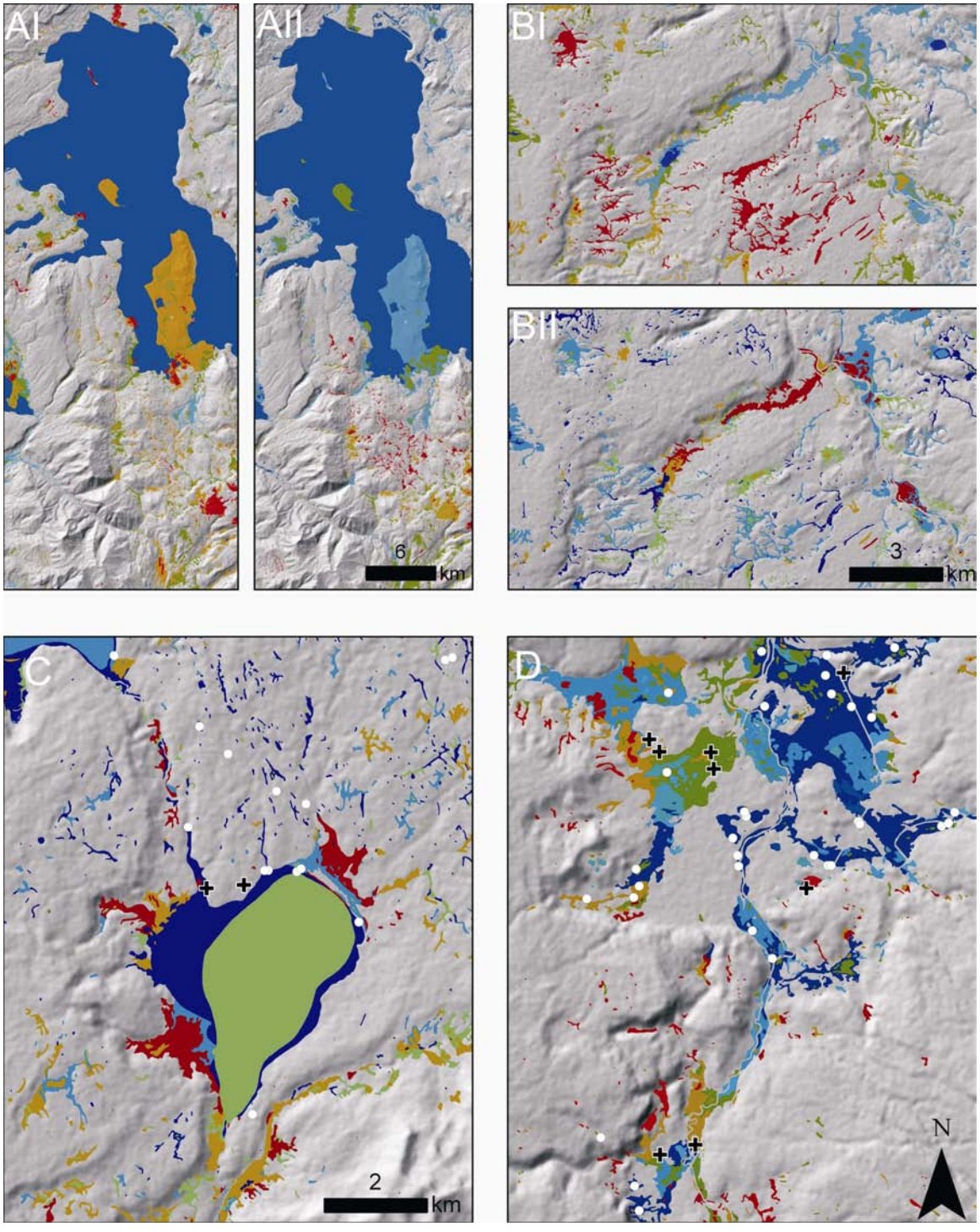


Figure A5.2



CHAPTER 3 - Quantifying *Bufo boreas* connectivity in Yellowstone National Park using landscape genetics

Melanie A. Murphy, Jeffrey S. Evans, Andrew Storfer

MA Murphy (corresponding author) (mamurphy@wsu.edu) School of Biological Sciences, Washington State University, Pullman WA, 99164. JS Evans USDA Forest Service, Rocky Mountain Research Station, 1221 S. Main, Moscow, Id 83843. A Storfer, School of Biological Sciences, Washington State University, Pullman WA, 99164.

In review with Ecology

ABSTRACT

A major objective of ecology is to understand how ecological processes limit dispersal and species' distributions. By spatially quantifying ecological components driving connectivity among occupied habitats, we can understand why some locally suitable habitats are unoccupied resulting in observed discontinuities in distribution. Using a novel application of spatial analysis and algorithmic modeling, we address the effects of three key ecological components on *Bufo boreas* genetic connectivity in Yellowstone National Park: ecological process, scale, and hierarchical organization. Although topographically corrected distance alone is significant, it explained little of the overall variation in genetic connectivity (10-21%). Habitat permeability, topographic morphology, and temperature-moisture are all significant processes influencing *B. boreas* connectivity. Connectivity was enhanced by

growing season precipitation, 1988 Yellowstone fires, cover, and length of frost free period. Conversely, connectivity was restricted by impervious surfaces (roads and development), hot-dry slopes, ridges, length of the hot season, and presence of complex topography (56% variation explained). In a hierarchical analysis, we were able to explain more variation within four genetic clusters as identified using STRUCTURE (74%; meadows, growing season precipitation, impervious surfaces) as opposed to between genetic clusters (45%; ridgelines, hot-dry slopes, length of hot season, and annual precipitation). We found that some metrics of ecological process operate on fine scales, while others operate at broad scales or across multiple scales making cross-scale spatial analysis critical for ecological interpretation. In the future, the approach we developed can be used to predict the impact of landscape change on *B. boreas* connectivity. Additionally, the analytical methods developed can be applied in any species or system with appropriate landscape and genetic data.

Keywords: Landscape genetics, ecoinformatics, multiple scales, RANDOM FORESTS, species connectivity, landscape ecology, algorithmic models, ecological process, Boreal toad

INTRODUCTION

A major objective of ecology is to understand how ecological processes that underlie dispersal affect species' distributions and population connectivity. Species' ranges are generally limited by biotic and abiotic factors (Morin et al. 2007). However, within a species range, underlying ecological processes influence connectivity among occupied habitat patches (Pulliam 2000, Ovaskainen and Hanski 2004). Quantifying spatial connectivity among these

habitats can elucidate why some locally suitable habitats are unoccupied, resulting in discontinuities of species distribution (Hanski and Gaggiotti 2004).

Amphibians generally have patchy distributions (Wagner and Fortin 2005) which may result in limited connectivity (Smith and Green 2005). Connectivity is likely driven by ecological processes (Turner 1989) such as habitat permeability (Mazerolle 2004), topographic morphology (Funk et al. 2005), and temperature-moisture regimes (Palo et al. 2003). Understanding how these ecological processes affect amphibian population connectivity and habitat occupancy is particularly important for conservation because many species are in decline for "enigmatic" reasons (Stuart et al. 2004).

Demographically derived measures of amphibian connectivity and distribution may vary drastically from year to year due to population stochasticity, possibly requiring decades of data for reliable estimates (Pechmann et al. 1991, Green 2003). However, connectivity among occupied habitats can be assessed by estimating genetic distance using neutral markers (Keyghobadi et al. 2005, Storfer et al. 2007), an effective approach for amphibian systems (Beebee 2005). Nonetheless, ecological data used to assess connectivity are often autocorrelated with non-linear relationships between dependent and independent variables, violating assumptions of standard parametric models (Wagner and Fortin 2005). In addition, these data and corresponding measures of genetic distance may have a low statistical signal to noise ratio resulting in models with little predictive power.

Ecoinformatic approaches (Chon and Park 2006), such as classification and regression trees, are well suited for studies of connectivity using genetic and ecological data. RANDOM FORESTS is an implementation of classification and regression trees that uses a bootstrap of the data to grow many trees (Breiman 2001, Cutler et al. 2007). This non-parametric approach

has the advantage of the ability to incorporate complex interactions among independent variables, measures of variable importance, and high power (Déath and Fabricius 2000, Cutler et al. 2007).

Boreal toads (*Bufo boreas*) in Yellowstone National Park are highly suitable for testing ecological processes driving connectivity as measured with neutral genetic data using an ecoinformatic approach. *Bufo boreas* is known to be sensitive to habitat permeability as measured by canopy cover (Bartelt et al. 2004), roads (Arens et al. 2007), and thermal influence (Koch and Peterson 1995). Topographically complex areas prevalent throughout Yellowstone may be barriers to gene flow (Loughheed et al. 1999) and temperature-moisture regimes likely limit connectivity (Palo et al. 2003, Bartelt and Peterson 2005). Each of these processes (habitat permeability, topographic morphology, and temperature-moisture) likely influences connectivity at multiple spatial scales. In addition, ecological processes likely operate in a hierarchy with fine-scale process influencing local connectivity and more broad-scale processes producing major breaks in genetic connectivity.

We test three hypotheses regarding *B. boreas* connectivity in Yellowstone National Park using neutral genetic markers: 1) connectivity of *B. boreas* breeding sites is a function of three ecological processes: habitat permeability, topographic morphology, and temperature-moisture regimes, 2) ecological processes in the surrounding landscape affect *B. boreas* connectivity across multiple spatial scales, and 3) connectivity is hierarchical with metrics operating at coarser spatial and temporal scales driving connectivity between genetic clusters; while metrics operating at finer spatial and temporal scales will drive connectivity within a genetic cluster.

METHODS

Field Data Collection

We stratified the National Wetlands Inventory data for Yellowstone National Park (USFWS 2002) by elevation and precipitation with at least twenty wetlands randomly sampled per stratum. To address low site occupancy (3%), we visited known *B. boreas* breeding sites in addition to the stratification to ensure adequate sampling (D. Patla, pub. comm.). Geographic site locations were recorded with a Garmin 12XL (Fig. 9). Tissue samples were collected by buccal swab from adults (Goldberg et al. 2003) and tail clip from tadpoles (2-5 mm) for DNA analysis following sterile field protocol.

Genetic Data Collection

DNA was extracted using the Qiagen DNeasy96 tissue protocol (Qiagen Inc). Fifteen microsatellite loci (Simandle et al. 2006) were genotyped with an ABI 3730 automated sequencer and scored with GeneMapper 3.7 (ABI) (for loci, optimized conditions, quality control, and basic validation see Appendix 6). A maximum likelihood algorithm (COLONY) was used to identify full sibling tadpoles (Wang 2004) and subsample sites at the size of the smallest sibling cluster to avoid biasing allele frequency distributions (C. Goldberg, pers. comm.). We represented the pair-wise genetic relationship between sample sites by connecting each site with a vector to every other site (Fig. 9B) in ArcInfo using a user designed Arc Macro Language (AML) program (ESRI 2000). We measured *B. boreas* connectivity for each vector with F_{ST} (Wright 1951, Weir and Cockerham 1984), D_{ps} (Bowcock et al. 1994), D_c (Cavelli-Sforza and Edwards 1967), and Nei's genetic distance (Nei 1978) as calculated in Microsatellite Analyser (MSA) (Dieringer and Schlötterer 2003).

Ecological Process

To evaluate the effect of each ecological process (habitat permeability, topographic morphology, temperature-moisture); we derived a set of metrics for each process from 30 m data (see Appendix 7 for detailed metric explanation, source data, predicted effect, and ecological justification). We assessed habitat permeability by cover class (forest, meadow, barren, water, wetland), percent cover, impervious surfaces, and thermal influence (Table 5; Appendix 7). We measured topographic morphology with a set of topographically derived variables including previously published (elevation, topographic roughness, elevation relief ratio, and ridges) and newly described metrics (relative slope position and hierarchical slope position (Table 5; Appendix 7). We evaluated temperature-moisture with metrics derived from the DEM and a spline-based climate model (Rehfeldt 2006, Rehfeldt et al. 2006) (Table 1). We calculated multiple temperature-moisture metrics based on topography: heat load index, compound topographic index, slope temperature-moisture, curvature, and solar insolation (Table 5; Appendix 7). Using the climate model, we calculated an average by vector for: growing season precipitation, mean annual precipitation, frost free period, and degree days >5° C (Table1). Due to the pair-wise nature of the data, we included topographic distance in all models.

We built a model of connectivity for each ecological process in RANDOM FORESTS (Breiman 2001, Liaw and Wiener 2002) and developed methodology for empirical model selection. We implemented RANDOM FORESTS in regression mode with 5000 trees, 34% out of bag (OOB) sample, and *m* (number independent variables iterated at each node) optimized to the OOB error estimate (Liaw and Wiener 2002). We removed multivariate redundant variables using qr matrix decomposition (threshold = 0.005) (Becker et al. 1988). In addition,

we developed methodology for empirical model selection in RANDOM FORESTS to improve model fit and ecological interpretability as follows. As an output of RANDOM FORESTS, metrics are ranked in order of importance based on the number of times a given metric improves overall model mean squared error (MSE). To compare variable importance among models, we developed a standardized measure (Model Improvement Ratio (MIR)). The MIR standardizes the mean increase in squared OOB residual error (MSE) importance measure to a ratio [MSE_n / MSE_{max}]. For each model we then subset variables > MIR threshold(s) (0.1-0.9 in 0.1 increments, replicates = 10). We then selected a model based on maximum improvement in MSE and percent variation explained, retaining all metrics above the MIR threshold.

We then assessed model fit by percent variation explained (pseudo-r^2), observed vs. predicted connectivity, residual distribution, MSE, and overall model significance. To assess overall model significance, we created a null distribution by randomizing observed genetic distances and calculating percent variation explained for each randomization (n=1000). We then calculated if the percent variation explained for a given model was > 95^{th} percentile of the null distribution (p-value < 0.05). RANDOM FORESTS iterates through independent variables at each node. Therefore, to evaluate the amount of variation explained by topographically corrected distance alone we needed to create random dummy metric in R (R development core 2007). Finally, we established the general nature of the relationship between connectivity and metrics of ecological process (+, - , =, complex) using a bivariate kernel density plot (Wand and Jones 1995).

Multiple scales

To assess the overall scale of landscape influence on *B. boreas* connectivity, we increased vector widths between sampling localities by creating round buffers around vectors in ArcInfo and recalculating all metrics (60, 120, 240, 480, and 960m diameter; ESRI 2000). For each scale, we tested the significance of each ecological process, created a final model using the above methods in RANDOM FORESTS and combined all selected metrics across all scales into one model to produce a cross-scale model of *B. boreas* connectivity. When removing multivariate redundant metrics, we retained the smallest scale if multiple scales of the same metric were statistically redundant as identified in the qr decomposition.

Hierarchical Effect

We partitioned the data into two hierarchical levels: vectors within a genetic cluster and vectors between genetic clusters. We identified genetic clusters using STRUCTURE (Pritchard et al. 2000) with 700,000 Markov-chain Monte-Carlo iterations (500,000 burn-in, 200,000 post burn-in) correlated allele frequencies, admixture model, inferred alpha with uniform prior, number of populations *(K)* 1-13, and 10 replicates. The number of genetic clusters *(K)* was estimated using ΔK (Evanno et al. 2005). To establish the validity of using identified genetic clusters as spatially defined groups and avoid the modifiable areal unit problem (MAUP) (Jelinski and Wu 1996), we assessed within versus among cluster genetic variation using an Analysis of Molecular Variance (AMOVA) (Peakall and Smouse 2006). For each hierarchical level, we tested significance of the three ecological processes and built a full model of connectivity using the above methods in RANDOM FORESTS and repeated the process at each scale (0, 60, 120, 240, 480, and 960 m). Using the metrics selected at each scale, we then built a cross-scale model for each hierarchical level, pruning metrics using MIR.

RESULTS

Genetic results

We collected *B. boreas* samples at 26 breeding locations for a total sample size of 805, reduced to 685 after screening tadpoles for full siblings (Fig. 9; see Appendix 6 for basic population genetic tests and statistics). Pair-wise F_{ST} ranged from 0 to 0.449; all pair-wise F_{ST} values were significantly greater than zero except between two sites around Alum creek (Fig. 9, square symbols). We were consistently able to explain more genetic variation using fewer metrics with the allele frequency distribution based measure D_{ps} than F_{ST}, D_{chord}, or D_n. Therefore, unless otherwise noted in the text, *B. boreas* connectivity is measured by the genetic distance statistic D_{ps} (0.1396 – 0.6437). Model results for all genetic distance measures are included in Appendix 8.

Ecological Process

Topographically corrected distance alone explained 21% of the variation in genetic distance ($p < 0.05$, Table 6). However, all tested ecological processes models were significant ($p < 0.05$, 52-55% variation explained) and explained more variation than distance alone (Table 6). The temperature-moisture models of connectivity explained more variation in genetic distance than habitat permeability or topographic complexity (55% vs. 52-54%, Table 6). Although the full model including all ecological processes explained only marginally more variation than any individual ecological process model (56%; Table 6), there was an overall improvement in model fit as assessed by MSE, observed vs. predicted plots, residual distribution, and model parsimony (Appendix 8). The full ecological process model included metrics of all the tested ecological processes: temperature-moisture (growing season

precipitation, slope-aspect transformation, degree days > 5° C (dd5), and frost free period), topographic morphology (topographic roughness, hierarchical slope position, elevation relief ratio), and habitat permeability (burned, canopy, and forest; Table 6).

Multiple Scales

We found that ecological processes and corresponding metrics operate at multiple spatial scales (Fig. 10). Overall, habitat permeability and topographic morphology do not show a clear trend between model performance and scale (Appendix 8). However, models of connectivity based on temperature-moisture improve with increasing scale (53.74 – 57.97%, Appendix 8). When all ecological processes are included in a single model, percent variation explained also improves with increasing scale (56.28% scale 0 – 59.99% scale 960 m; Appendix 8). In addition, metrics of these ecological processes had distinct scale effects around 240 m (Fig. 10). Two metrics, growing season precipitation and slope temperature-moisture were important across all scales (Fig. 10; 11). Additional temperature-moisture (dd5 and frost free period), habitat permeability (burned, canopy), and topographic morphology metrics (roughness, elevation relief ratio and hierarchical slope position) function at ≤ 240 m (Fig. 10). Conversely, ridge lines are only an important metric ≥ 240 m (Fig. 10). We were able to explain the most variation in genetic distance by creating a single cross scale model (59.09%, Table 6). The cross scale model had an overall improvement in model fit assessed by MSE, observed vs. predicted plots, residual distribution, and model parsimony (Appendix 8). The cross scale model included: topographic morphology (ridges - 960 m, elevation relief ratio - 60 m), temperature-moisture regime (growing season precipitation and hot-dry slopes), and habitat permeability (canopy -60 m, impervious surfaces - 240 m; Table 6).

Hierarchical Effect

We identified 4 well supported genetic clusters using STRUCTURE, three of which contained substructure (Fig. 9) (Pritchard et al. 2000, Evanno et al. 2005). The four main genetic clusters are spatially justifiable units; they minimize within cluster variation and maximize between cluster genetic variation as established with an AMOVA ($p < 0.001$). As predicted, different ecological processes explain within cluster variation compared to between cluster variation (Table 6). In addition, at a given hierarchical level (within or between genetic clusters) distance was not selected as an important variable. Within clusters, we were able to explain 74.03% of the variation in D_{ps}, the most variation explained in any tested model, with little model error (Table 6). Meadows, growing season precipitation (at multiple scales), and impervious surfaces are all important variables (Fig. 11; Table 6). Between clusters, we were able to explain 45.32% of the variation in genetic distance (Table 6) with F_{ST} and D_{ps} models of connectivity explaining an equivalent amount of variation (45.32% D_{ps} vs. 44.85% F_{ST}, Table 6). Between clusters, connectivity is negatively associated with hot-dry slopes (ssina), dd5, and ridges (Fig. 11f,h,i). Connectivity is positively associated with mean annual precipitation and canopy cover (Fig. 11c; Table 6).

DISCUSSION

We contribute to ecological understanding of species fine-scale distributions by explaining the influence of key ecological processes on population connectivity utilizing population genetic data. Habitat permeability, topographic morphology, and temperature-moisture regimes are all strongly correlated with *B. boreas* connectivity in Yellowstone, with relatively little variation explained by distance alone. Processes driving connectivity operate

at multiple spatial scales. Habitat permeability tends to operate on fine scales, topographic morphology at broad scales, and temperature-moisture across spatial scales. Hierarchical structure is present, with finer spatial and temporal processes driving connectivity within genetic clusters while broader spatial and temporal processes drive connectivity between genetic clusters. Finally, this study puts the Yellowstone distribution of *B. boreas* in an ecological context. Distribution may be limited by growing season precipitation, ridges, and impervious surfaces resulting in the observed patchy distribution (Table 6).

Genetic Metrics

Our results demonstrate that allele frequency distribution based genetic distance (D_{ps}) is more powerful and detects more recent landscape changes than heterozygosity based genetic distance (F_{ST}). Allele frequency distribution based methods are more likely to detect recent events (Miller and Waits 2003), identifying landscape genetic structure within five generations post change (Murphy et al. in revisions). Conversely, F_{ST} may reflect past landscape condition (Holzhauer et al. 2006) and assumes population equilibrium which is often violated in natural populations.

Ecological Process

In contrast to other studies (Arnaud 2003, Manier and Arnold 2006), topographic distance was not the strongest predictor of connectivity (Table 6; Appendix 8). Instead, habitat permeability, topographic morphology, and temperature-moisture all significantly influence *B. boreas* connectivity and are represented in the final model, although temperature-moisture regime had the strongest influence (Table 6).

The importance of habitat permeability, topographic morphology, and temperature-moisture for *B. boreas* connectivity is well supported in the literature. Habitats with mid-

range canopy cover (such as shrubby areas) are preferentially selected during dispersal, thus increasing connectivity (Bartelt et al. 2004). Post-fire habitats also result in increased connectivity through *B. boreas* colonization of new habitats as well as increased dispersal in the years immediately following fire (Hossack and Corn 2007). This is possibly due to terrestrial microhabitats provided by woody debris, increased water temperature at breeding sites, and stable temperature of burrowing sites in burned areas (Hossack and Corn 2007). A study of tiger salamanders in northern Yellowstone also found a positive correlation between burned habitats and gene flow (Spear et al. 2005). Measures of topographic morphology represent physiological cost of moving through the landscape and across watershed boundaries, which may be barriers for amphibians (Funk et al. 1999). Growing season precipitation corresponds to the *B. boreas* active period, indicating breeding site persistence (Corn 2003) and habitat moisture. Metamorphosed *B. boreas* have little physiological control over evaporation (Bartelt and Peterson 2005); therefore, individuals likely prevent desiccation by avoiding hot, exposed habitats. Finally, length of the growing season (frost free period) affects time to metamorphosis in amphibians and may limit gene flow among populations (Giordano et al. 2007).

Multiple Scales

We demonstrate that metrics of ecological processes influencing *B. boreas* connectivity operate at multiple scales. Fine-scale metrics include measures of local habitat condition, microclimate, and topography in which local variability influences *B. boreas* connectivity (Table 6). Two key metrics, growing season precipitation and hot-dry slopes, operate across scales suggesting a more general relationship with *B. boreas* connectivity (Fig. 10).

Each scale, analyzed independently, results in a slightly different inference about *B. boreas* connectivity. Incorporating all of these scales into a single analysis captures a more complete picture of the effects of ecological processes on *B. boreas*. The cross scale model includes metrics measured at fine scales (canopy, elevation relief ratio), mid-scale (impervious surfaces), broad scale (ridges), and across multiple scales (growing season precipitation and hot-dry slopes; Fig. 10). However, cross-scale models only improved the percent variation explained by ~3-5%, indicating that the straight-line model (scale 0) had high predictive power. This model of connectivity may be performing well for two major reasons: inclusion of multi-scale metrics and the protected nature of the study system. Multi-scale metrics (i.e., multiple window sizes) may capture the surrounding landscape, even at finer-scales. In addition, although Yellowstone has areas of high human use, it is a relatively intact landscape. For this reason, historic and current gene flow may be very similar strengthening the overall observed pattern of connectivity. The influence of the surrounding landscape and subsequent scale effect may be more notable in highly fragmented systems.

Hierarchical Effect

Within cluster variation in connectivity represents a finer spatial and temporal scale than between cluster variation. Between clusters, genetic variation represents historic or long-term conditions. Therefore, it is not surprising that we are able to explain more variation in connectivity within clusters (74.03%) as compared to between clusters (44.85%, Table 6) using current landscape condition. Topographic distance is notably absent as an important variable in predicting *B. boreas* connectivity both within and between genetic clusters. These results reinforce the cross-scale analyses and suggest two things: 1) distance may be acting as

a surrogate for the hierarchical effect in the ecological process and scale models and 2) landscape features can have a strong effect on connectivity independent of distance.

Overall, within genetic cluster connectivity is governed by habitat permeability and moisture; areas with high growing season precipitation, mid-range cover and low impervious surface density are positively associated with high connectivity (Table 6). The overwhelming importance of growing season precipitation within genetic clusters may indicate fragmentation due to drought conditions over the past 8-12 years (Table 6). These effects are also consistent with apparent drought-induced genetic bottlenecks in tiger salamanders throughout northern Yellowstone (Spear et al. 2006). In addition, growing season precipitation is important at multiple non-redundant scales suggesting it represents multiple mechanisms. Local growing season precipitation may represent average humidity and soil moisture. At the 960m scale, growing season precipitation may represent regional differences in temperature-moisture regime across Yellowstone.

Connectivity between genetic clusters can be explained by more general, mostly broad scale metrics. Major topographic features, such as ridgelines, can explain the observed genetic groups suggesting that topography is a barrier for *B. boreas* in Yellowstone. In contrast to the multi-scale models that are based on all pair-wise connections, hot-dry slopes only operate at mid to broad scales and degree days > 5° C only at broad scales (Table 6). Mean annual precipitation replaces growing season precipitation as the prominent measure of moisture across scales, reinforcing the more long-term nature of processes driving connectivity between genetic groups.

CONCLUSIONS

Our study has several important implications for ecological applications of genetic data, *B. boreas* conservation, and understanding mechanisms of fine-scale species distributions. Isolation-by-distance and autocorrelation are important considerations in landscape genetics. However, even in relatively intact landscapes such as Yellowstone, distance may only explain a small portion of the variation in genetic connectivity compared to the variation explained in models including other potential explanatory processes. By using an algorithmic approach, we are able to simultaneously include multiple metrics of ecological processes to explain a high proportion of variation in genetic distance (up to 74%).

Because *B. boreas* is thought to be in decline throughout large portions of its range (Muths et al. 2003), it is essential to assess species' connectivity and vulnerability in protected areas where populations remain. At all scales (Table 6; Appendix 8), precipitation and temperature are major drivers of *B. boreas* connectivity. As Yellowstone becomes warmer and drier as predicted due to climate change (Bartlein et al. 1997), connectivity in this landscape will likely degrade. In addition, niche space for species limited by temperature-moisture regimes will likely shift (Rehfeldt et al. 2006), potentially resulting in much of the presently occupied habitat being unsuitable. Knowledge of these threats may help inform management decisions; e.g., an increase in impervious surfaces due to development, also shown to limit this species' connectivity, may have a greater impact in a context of climate change than it would otherwise. Finally, implementation of landscape genetic approaches can provide robust estimates of population connectivity to assist in development of science-based conservation and management decisions.

ACKNOWLEDGEMENTS

Research was permitted by Yellowstone National Park (Permit YELL-05452**;** C Smith and C Hendrix) and IACUC (ASAF #3378). We would like thank M Murphy, M Flisher, D Murphy for field assistance; C Corkan, D Corkan, D Patla and C Peterson provided additional tissue samples. A Gallant, P Bartelt, and R Klaver assisted in compiling spatial data. S Cushman, R Dezzani, C Goldberg, D Patla, C Peterson, D Pilliod, S Spear, M Webster and L Waits provided valuable comments and advice. Funding for MAM was provided by EPA-STAR fellowship (FP-916695**)**, NSF (grant DEB-0608458 to MAM and AS**)**, James King fellowship, Theodore Roosevelt Memorial Fund, Society for Wetland Science (2004, 2005), WSU Zoology alumni scholarship, Canon National Parks, Sigma-Xi grants-in-aid (2004), and Graduate Women in Science. AS was supported by NSF DEB-0548415.

LITERATURE CITED

Arens, P., T. van der Sluis, W. P. van't Westende, B. Vosman, C. C. Vos, and M. J. M. Smulders. 2007. Genetic population differentiation and connectivity among fragmented Moor frog (*Rana arvalis*) populations in The Netherlands. Landscape Ecology 22:1489-1500.

Arnaud, J.-F. 2003. Metapopulation genetic structure and migration pathways in the land snail *Heliz aspersa*: influence of landscape heterogeneity. Landscape Ecology 18:333-346.

Bartelt, P., and C. R. Peterson. 2005. Physical modeling operative temperature and evaporative water loss in amphibians. Journal of Thermal Biology 30:93-102.

Bartelt, P., C. R. Peterson, and R. W. Klaver. 2004. Sexual differences in the movements and habitats selected by Western toads in southeastern Idaho. Herpetologica 60:455-467.

Bartlein, P. J., C. Whitlock, and S. L. Shafer. 1997. Future climate in the Yellowstone National Park region and its potential impact on vegetation. Conservation Biology 11:782-792.

Becker, R., J. Chamers, and A. Wilks. 1988. The new S language: a programming environment for data analysis and graphics. Murry Hill, New Jersey.

Beebee, T. J. C. 2005. Conservation genetics of amphibians. 95:423-427.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455-457.

Breiman, L. 2001. Random Forests. Machine Learning 45:5-32.

Cavelli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. American Journal of Human Genetics 19:233-257.

Chon, T.-S., and V.-S. Park. 2006. Ecological informatics as an advanced interdisciplinary interpretation of ecosystems. Ecological Informatics 1:213-217.

Corn, P. S. 2003. Amphibian breeding and climate change: importance of snow in the mountains. Conservation Biology 17:622-625.

Cutler, D. R., T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

Déath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178-3192.

Dieringer, D., and C. Schlötterer. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3:167-169.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620.

Evans, I. S. 1972. General geomorphometry, detivatives of altitude, and descriptive statistics. *in* R. J. Chorley, editor. Spatial Analysis in geomorphology. Harper & Row, New York.

Fu, P., and P. M. Rich. 1999. Design and implementation of the Solar Analyst: an ArcView extension for modeling solar radiation at landscape scales. *in* 19th Annual ESRI User Conference. ESRI, San Diego, CA.

Funk, W. C., M. S. Blouin, P. S. Corn, B. A. Maxell, D. S. Pilliod, S. Amish, and F. W. Allendorf. 2005. Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. Molecular Ecology 14:483-496.

Funk, W. C., D. A. Tallmon, and F. W. Allendorf. 1999. Small effective population size in the long-toed salamander. Molecular Ecology 8:1633-1640.

Giordano, A. R., B. J. Ridenhour, and A. Storfer. 2007. The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactylum*). Molecular Ecology 16:1625-1637.

Goldberg, C. S., M. E. Kaplan, and C. R. Schwable. 2003. From the frog's mouth: buccal swabs for collection of DNA from amphibians. Herpetological Review 34:220-221.

Green, D. M. 2003. The ecology of extinction: population fluctuation and decline in amphibians. Biological Conservation 111:331-343.

Hanski, I., and O. E. Gaggiotti, editors. 2004. Ecology, genetics, and evolution of metapopulations. Elsevier, Burlington MA.

Holzhauer, S., K. Ekschmitt, A.-C. Sander, J. Dauber, and V. Wolters. 2006. Effect of historic landscape change on the genetic structure of the bush-cricket *Metrioptera roeseli*. Landscape Ecology 21:891-899.

Hossack, B. R., and P. S. Corn. 2007. Responses of pond-breeding amphibians to wildfire: short-term patterns in occupancy and colonization. Ecological Applications 17:1403-1410.

Jelinski, D. E., and J. Wu. 1996. The modifiable areal unit problem and implications for landscape ecology. Landscape Ecology 11:129-140.

Keyghobadi, N., J. Roland, S. F. Matter, and C. Strobeck. 2005. Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proceedings of the Royal Society of London 272:553-560.

Koch, E. D., and C. R. Peterson. 1995. Amphibians and Reptiles of Yellowstone and Grand Teton National Parks. University of Utah Press, Salt Lake City.

Liaw, A., and M. Wiener. 2002. Classification and regression by Random-Forest. R News 2:18-22.

Lougheed, S. C., C. Gascon, D. A. Jones, J. P. Bogard, and P. T. Boag. 1999. Ridges and rivers: a test of competing hypotheses of Amazonian diversification using a dart-poison frog (Epipedobates femoralis). Proceedings of the Royal Society B 266:1829-1835.

Manier, M. K., and S. J. Arnold. 2006. Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. Proceedings of the Royal Society B 273:3001-3009.

Mazerolle, M. J. 2004. Amphibian road mortality in response to nightly variations in traffic intensity. Herpetologica 60:45-53.

McCune, B., and D. Keon. 2002. Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science 13:603-606.

McGarigal, K., and B. J. Marks. 1995. FRAGSTATS. Spatial analysis program for quantifying landscape structure. USDA Forest Service General Technical Report PNW-GTR-351.

Miller, C. R., and L. P. Waits. 2003. The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): Implications for conservation. Proceedings of the National Academy of Science 100:4334-4339.

Moore, I., P. Gessler, G. Nielsen, and G. Petersen, editors. 1993. Terrain attributes and estimation methods and scale effects. Wiley, London.

Morin, X., C. Augspurger, and I. Chuine. 2007. Process-based modeling of species' distributions: what limits temperate tree species' range boundaries? Ecology 88:2280-2291.

Murphy, M. A., J. S. Evans, S. Cushman, and A. Storfer. in revisions. Evaluation of a novel approach for representing "populations" as continuous surfaces in landscape genetics. Ecography.

Muths, E., P. S. Corn, A. P. Pessier, and D. E. Green. 2003. Evidence for disease-related amphibian decline in Colorado. Biological Conservation 110:357-365.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Ovaskainen, O., and I. Hanski. 2004. Metapopulation dynamics in highly fragmented landscapes. Pages 73-104 *in* I. Hanski and O. E. Gaggiotti, editors. Ecology, genetics, and evolution of metapopulations. Elsevier, Burlington MA.

Palo, J. U., R. B. O'Hara, A. T. Laugen, A. Laurila, C. R. Primmer, and J. Merila. 2003. Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. Molecular Ecology 12:1963–1978.

Peakall, R., and P. E. Smouse. 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Pechmann, J. H. K., D. E. Scott, R. D. Semlitsch, J. P. Caldwell, L. J. Vitt, and J. W. Gibbons. 1991. Declining amphibian populations: the problem of separating human impacts from natural fluctuations. Science 253:892-895.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pulliam, H. R. 2000. On the relationships between niche and distribution. Ecology Letters 3:349-361.

Rehfeldt, G. E. 2006. A spline model of climate for the Western United States. RMRS-GTR-165, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Rehfeldt, G. E., N. L. Crookston, M. V. Warwell, and J. S. Evans. 2006. Empirical analysis of plant-climate relationships for western United States. International journal of plant sciences 167:1123-1150.

Riley, S. J., S. D. DeGloria, and R. Elliot. 1999. A terrain ruggedness index that quantifies topographic heterogeniety. Intermountain Journal of Sciences 5:23-27.

Simandle, E. T., M. M. Peacock, L. Zirelli, and C. R. Tracy. 2006. Sixteen microsatellite loci for the Bufo boreas group. Molecular Ecology Notes 6:116-119.

Smith, A. M., and D. M. Green. 2005. Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations? Ecography 28:110-128.

Spear, S., C. R. Peterson, M. Matacq, and A. Storfer. 2006. Molecular evidence for historical and recent population size reductions of tiger salamanders (*Ambystoma tigrinum*) in Yellowstone National Park. Conservation Genetics 7:605-611.

Spear, S. F., C. R. Peterson, M. Matacq, and A. Storfer. 2005. Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). Molecular Ecology 14:2553-2564.

Stage, A. 1976. An expression of the effects of aspect, slope, and habitat type on tree growth. Forest Sciences 22:457-469.

Storfer, A., M. A. Murphy, J. S. Evans, C. Goldberg, S. Spear, S. Robinson, R. Dezzani, E. Demmelle, L. Vierling, and L. P. Waits. 2007. Putting the 'landscape' in landscape genetics. Heredity 98:128-142.

Stuart, S. N., J. S. Chanson, N. A. Cox, B. E. Young, A. S. L. Rodrigues, D. L. Fischman, and R. W. Waller. 2004. Status and trends of amphibian declines and extinctions worldwide. Science 306:1783-1786.

Turner, M. G. 1989. Landscape Ecology: the effect of pattern on process. Annual Review of Ecology and Systematics 20:171-197.

USFWS. 2002. National Wetlands Inventory.

Wagner, H. H., and M.-J. Fortin. 2005. Spatial analysis of landscapes: concepts and statistics. Ecology 86:1975-1987.

Wand, M. P., and M. C. Jones. 1995. Kernal Smoothing. Chapman & Hall, London.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics 166:1963-1979.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

Zeverbergen, L. W., and C. R. Thorne. 1987. Quantitative analysis of land surface topography. Earth Surface Processes and Landforms 12:47-56.

TABLES

Table 5

Metrics for each ecological process, code, and data source or metric derivation. Source acronyms: SRTM (Space Radar Telemetry Mission), DEM (Digital Elevation Model), NLCD (2001 National Land Cover Database)Table is expanded to include predicted relationship, metric explanation, and ecological justification in Appendix 7.1. In addition, definitions for newly derived metrics (rsp, hsp) are included in Appendix 7.2.

Process	**Metric**	**Code**	**Source**
IBD	Topographic Distance (m)	distance	SRTM DEM
Habitat	Water	water	NLCD: 11
	Forest	forest	NLCD: 41, 42, 43
	Wetlands	wetland	NLCD: 90, 95
	Meadow	meadow	NLCD: 71, 72
	Canopy Percent	manopy	% cells classified as forest
	Impervious surfaces	imperv	NLCD: 21, 22, 23, 24
	Thermal[1]	therm	(McGarigal and Marks 1995)
	1988 Fire perimeter	burn	USGS
Topo-morph	Elevation (m)	elev	SRTM DEM
	Topographic roughness[2]	rough	(Riley et al. 1999)
	Elevation relief ratio err[2]	err	(Evans 1972)
	Relative slope position[3]	rsp	Newly derived (Appendix 7.2)
	Hierarchical slope position[3]	hsp	Newly derived (Appendix 7.2)

	Ridges	ridge	Hsp threshold
Temp-Moist	Heat Load Index	hli	(McCune and Keon 2002)
	Compound topographic index	cti	(Moore et al. 1993)
	Slope-aspect transformation	ssina	(Stage 1976)
	Curvature	crv	(Zeverbergen and Thorne 1987)
	Solar insolation	inso	(Fu and Rich 1999)
	Growing season precip	gsp	(Rehfeldt 2006)
	Mean annual precip	map	(Rehfeldt 2006)
	Frost free period	ffp	(Rehfeldt 2006)
	Degree days > 5° C	dd5	(Rehfeldt 2006)

Some metrics were calculated at multiple window sizes [1] Percent of landscape (PLAND) thermal at 0, 13, and 26 cell radii. [2]Calcualted at 3X3, 15X15, and 27X27 cell window sizes. [3]Decomposed from 0 – 27X27 cell window size (see Appendix 7.2).

Table 6

Landscape genetic models of B. boreas connectivity in Yellowstone National Park. Group is the group tested within a hypothesis. Distance is genetic distance as F_{ST} or D_{ps}. PVE is percent variation explained, a pseudo-R^2 from RANDOM FORESTS. Model is the selected model with metrics in descending order of importance. Type indicates type of relationship: **bold –** negative relationship, normal – positive relationship, and *italics* – no single direction across the range of the metric. Numbers proceeding a _ indicated window size for multi-scale variables and numbers following a _ indicate buffer size. For full model results and additional validation statistics, see Appendix 8.

Hypothesis	Group	Distance	PVE	Model
Distance	Distance	FST	10.050	**Distance**
		DPS	21.069	**Distance**
Process	Habitat	FST	33.481	Meadow**,** *canopy*, wetland**, distance,** burn**,** *forest*, therm720**, imperv**
		DPS	51.797	burn, **distance**, *canopy*, *water***,** *forest*, meadow, **imperv,** wetland, therm360
	Topo	FST	34.357	*elev*, **distance**, **rough15**, *rsp*, **rough3**, **rough27**, **err27**, *hsp*, **err15**, **err3**
		DPS	53.570	**Hsp**, *elev*, **rough15**, **err27**, **distance**,**rough27**, **err3**, **err15, rough3**
	Temp-moist	FST	39.939	map, gsp, **distance**, **hli, dd5,** ffp
		DPS	55.309	Gsp, **distance**
	All	FST	38.420	meadow, wetland, gsp, map, *canopy*, **dd5**, *elev*, **rough3**, **rough15**, ffp, burn, **errr27**, **imperv**, **distance**, *rsp*, **err15**, **hli**, **rough27**, therm720, *forest*, *hsp*
		DPS	56.284	gsp, **distance**, ssina, **rough27/15**, **burn, canopy, hsp**, **err27, dd5, err15, forest, ffp**
Scale	X-Scale	FST	43.925	Meadow_60,meadow_480,*imperv_480*,meadow_240,**ridge_960**, *canopy*, meadow_960 *imperv_240*, **err15_60, ridge_240, dd5_960, err15_120,** wetland_960, gsp_60, **dd5_60, imperv_120,** map_120, **imperv_60,** *canopy_120***, imperv_960**, gsp_120, gsp_480
		DPS	59.087	**ridge_960**, gsp, **ssina_480,** *canopy_60*, **ssina_240, ssina_960, imperv_240, err27_60**
Hierarchical	Within	FST	38.902	**imperv_960**, *err27*, *canopy*
		DPS	74.027	Meadow_480, gsp_960**, imperv_480,** gsp_120, gsp_480, gsp
	Between	FST	45.324	*canopy_960*, **ridge_480, dd5_960, ssina_480,** gsp_240, meadow_60, **ridge_240,** meadow_960, **ssina_240,** wetland_960, meadow_480, **ridge_960,** gsp_480, **imperv_240,** meadow_120, gsp_960, **imperv_120, err15_60,** *water_120*, **imperv_60,** meadow_240, *canopy_120*, *canopy_60*, **err15_120,** cti_120, cti_60, *canopy*, **imperv_480,** gsp_120, gsp_60
		DPS	44.848	**ssina_240, ssina_480,** map, **ssina_960, ridge_960,** *canopy_480*, map_240, meadow_960,*canopy_240*,**imperv_240**,map_120,**imperv_480**,**imperv_120, dd5_960**

FIGURE LEGENDS

Figure 9. A) Study area (Yellowstone National Park) with sample locations and genetic groups as identified in STRUCTURE. We identified four levels of genetic structure (circles, stars, triangles, and squares) with substructure in three of those groups (circles, starts, triangles) displayed by symbol shade (white, gray, or black). Numbers indicate areas where more than one site was present (see blow-up of these areas in Appendix 9). **B)** Pair-wise connections between sites overlaid on a shaded relief. Figure can be seen in color in Appendix 9.

Figure 10. Scale of ecological process (m) for selected metrics (metric abbreviations are given in Table 5). Bars indicate at which scales a metric was selected, bar color indicates ecological process.

Figure 11. Select important metrics mapped for the study area by ecological process (habitat permeability (A-C), ecological morphology (D-F), and temperature-moisture (G-I)). Circles represent sample locations. Metrics are as follows: A) 1988 fire perimeters (gray) with impervious surfaces overlaid, B) NLCD classification showing water (white), open/shrub (light gray), forest (dark gray), and wetlands (black), C) canopy cover (0 (white)- 100% (black)), D) roughness at 15X15 window (0 (white)- 100% (black)), E) elevation relief ratio (0 (white)- 100% (black)), F) ridges (black), G) growing season precipitation (low (white)- high (black)), H) ssina (0 (white) – 100% (black)), and I) dd5 (low (white) – high (black)). For full color see Appendix 9.

FIGURES

Figure 9

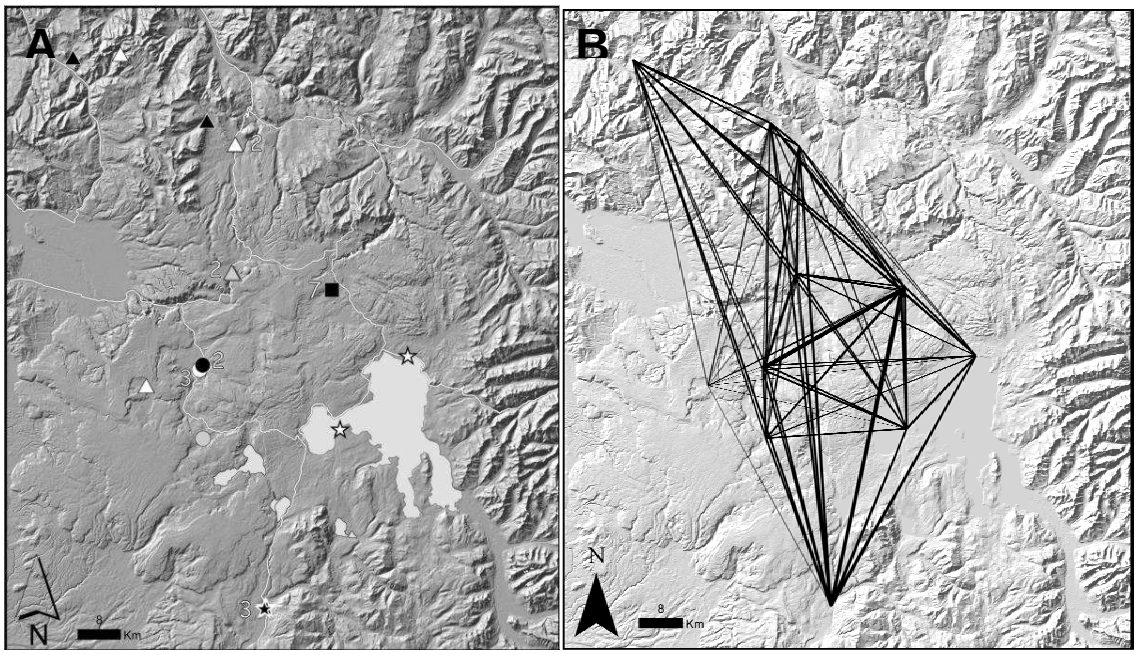


Figure 10

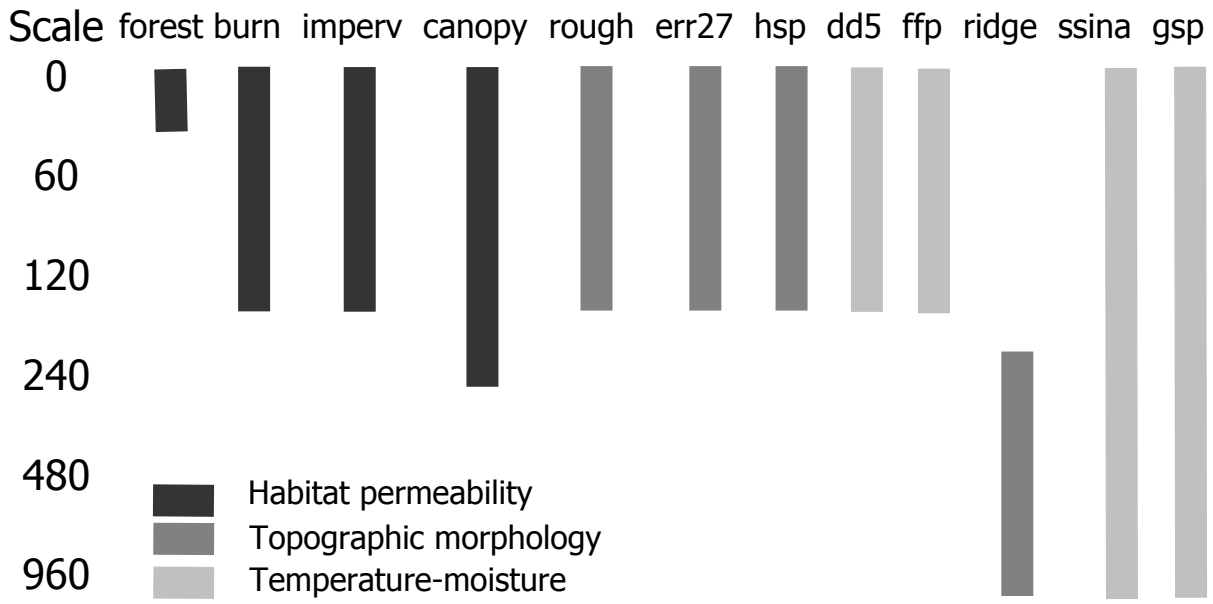
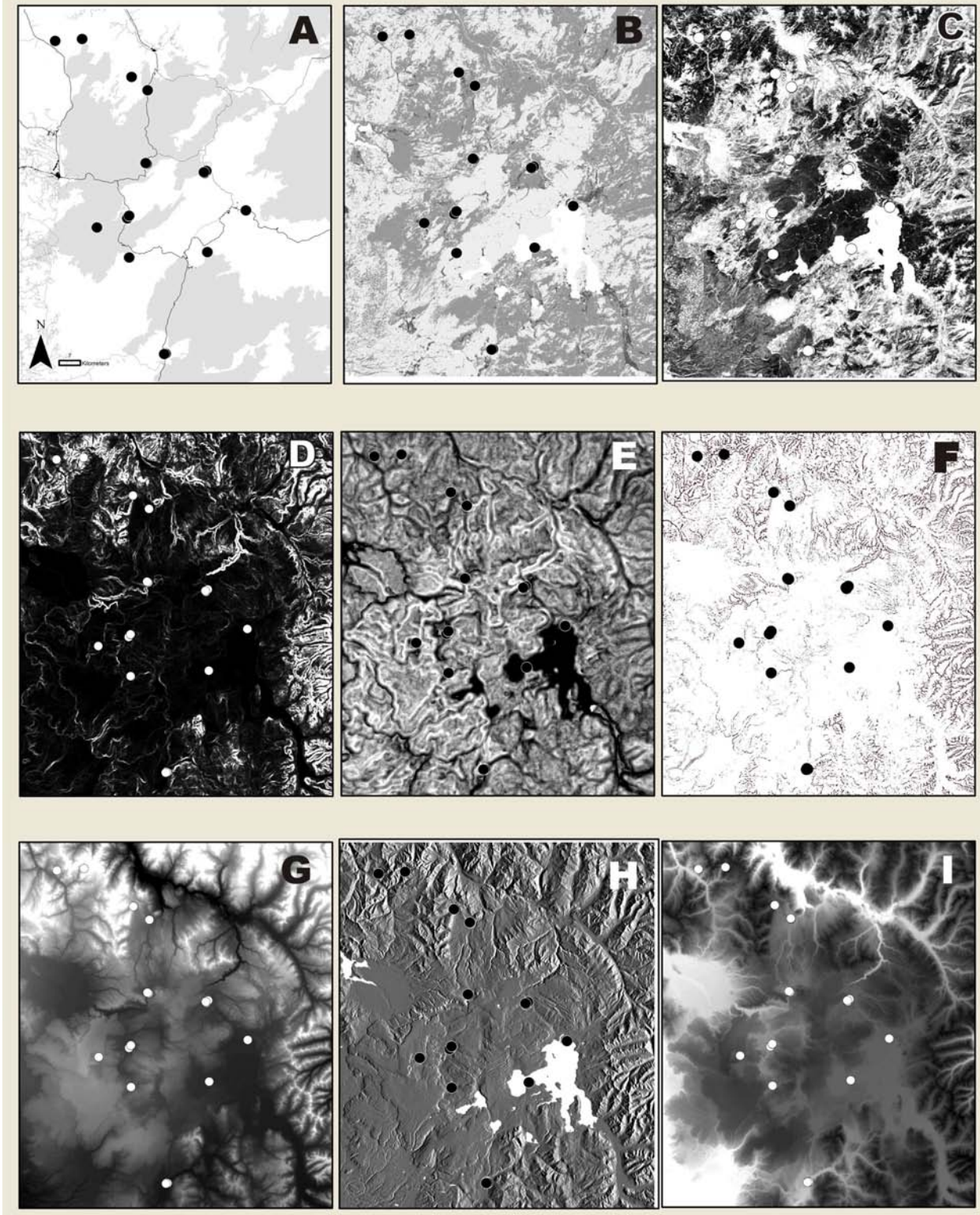

Scale
forest
burn
imperv
canopy
rough
err27
hsp
dd5
ffp
ridge
ssina
gsp
0
60
120
240
480
960
Habitat permeability
Topographic morphology
Temperature-moisture

Figure 11



Appendix 6.

Appendix 6.1 Genetic summary information. To ensure genotype accuracy, we included at least 2 negative controls per extraction and PCR, amplified a known genotype in each reaction, and reamplified all rare alleles (<5% frequency) in addition to at least 10% of samples from each PCR to screen for genotyping and human error. After subsampling sibling clusters (Wang 2004, C. Goldberg pers. comm.), all loci were tested for global concordance with Hardy-Weinberg equilibrium, significant linkage disequilibrium between all pairs of loci, and Hardy-Weinberg equilibrium at each site using GENEPOP and MICROSATELLITE ANALYSER (MSA) (Raymond and Rousset 1995, Dieringer and Schlötterer 2003). No loci or populations were out of Hardy-Weinberg equilibrium or linkage equilibrium (Raymond and Rousset 1995) with the exception of three sites with small sample sizes which were excluded from the pair-wise analysis (see Fig3.1A, B).

Appendix 6.2 Microsatellite conditions and basic statistics. Includes the following: locus (Simandle et al. 2006), PCR multiplex (what primers were in a single PCR reaction), panel (what product were run in a single capillary), fluorescent label color (R- red, G- green, B- blue, Y-yellow), primer concentration in PCR reaction, number of alleles found, and expected heterozygosity. Qiagen multiplex kit (including Q-solution) was used for all reactions with standard conditions. General PCR cycle was: initial denaturing (95° C) 15 min, 35 cycles (94° C 30 sec denature, X° C annealing 90 sec, 72° C extension 60 sec), and 60° C 60 min final extension.

Locus	**Multiplex**	**Panel**	**Label**	**[Primer]**	**Annealing**	**# alleles**	**H_{exp}**
BBR36	1	1	R	.31 μM	56 °C	11	0.8261
BBR233	1	1	G	.31 μM	56 °C	6	0.2768
BBR29	1	1	B	.26 μM	56 °C	3	0.3227
BBR86	1	1	B	.26 μM	56 °C	7	0.7910
BBR87b	1	1	G	.26 μM	56 °C	16	0.9041
BBR17	1	1	Y	.26 μM	56 °C	2	0.1893
BBR297	1	1	G	.26 μM	56 °C	10	0.6006
BBR4-2	1	1	B	.26 μM	56 °C	4	0.1482
BBR281	2	2	R	.26 μM	55 °C	10	0.4522
BBR292	2	2	B	.26 μM	55 °C	14	0.8093
BBR16	2	2	B	.26 μM	55 °C	22	0.9223
BBR4	2	2	G	.26 μM	55 °C	5	0.6414
BBR293	2	2	B	.26 μM	55 °C	10	0.2855
BBR34-2	2	2	R	.31 μM	55 °C	12	0.8224
BBR45	single	1	Y	.26 μM	53 °C	15	0.7426
Average						**9.8**	**0.5823**

Literature Cited

Dieringer, D., and C. Schlötterer. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3:167-169.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Raymond, M., and F. Rousset. 1995. GENEPOP(1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248-249.

Simandle, E. T., M. M. Peacock, L. Zirelli, and C. R. Tracy. 2006. Sixteen microsatellite loci for the Bufo boreas group. Molecular Ecology Notes 6:116-119.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics 166:1963-1979.

Appendix 7

Metric explanation, source data, predicted effect, ecological justification, and new metric definitions.

Appendix 7.1 Table of ecological processes and respective metrics. Process – process of which the metric is a measure (distance (IBD), habitat permeability (Habitat), topographic morphology (Topo), or temperature-moisture (Temp-Moist). Metric – metric name. Code- metric code or abbreviation. Source – source of data which contained the variable or from which the variable was derived. Sources are as follows: NLCD – National Land Cover Database (2001), USGS – coverages developed for Yellowstone National Park research unit, SRTM - Shuttle Topographic Radar Mission digital elevation model (DEM), and Rehfeldt 2006 – climate spline model that adjusts weather stations using elevation (used SRTM data). Predicted relationship (Pred) – if we expected to find a positive (+), negative (-), or neutral (=) relationship between the metric and connectivity. Metric explanation – brief description of variable and reference if a previously published metric. Calculation – description of how the metric was calculated or reference. Ecological justification – brief justification for including metric in analysis with select reference(s).

Process	Metric	Code	Source	Pred	Metric Explanation	Calculation	Ecological Justification
IBD	Topographic Distance (m)	distance	SRTM	-	Topographically corrected distance	Sum	Isolation-by-distance (IBD) has been seen in *B. boreas* (Manier and Arnold 2006)
Habitat	Water	water	NLCD	-	Class Open water	Count of cells	Large areas of open water likely barrier to dispersal
	Forest	forest	NLCD	+	Classes Deciduous, Evergreen, Mix	Count of cells	Forest provides cover (Bartelt et al. 2004)
	Wetlands	wetland	NLCD	+	Classes Emergent & Herbaceous	Count of cells	Provide food, breeding, and moisture
	Meadow	meadow	NLCD	+	Classes: 71, 81, 82	Count of cells	Provide shrub cover (Bartelt et al. 2004)
	Barren	barren	NLCD	-	Classes: 31, 12	Count of cells	Little cover, exposed (Bartelt et al. 2004)

	Canopy Percent	canopy	NLCD	+	Percent cells forest cover	# cells forest/total # cells	Use mid-range cover (Bartelt et al. 2004)
	Impervious surfaces	imperv	NLCD	-	0-100% impervious surface	Count of cells	Roads & development may limit dispersal ability (Arens et al. 2007)
	Thermal[1]	therm	USGS	-, +, =	Percent thermal cells	PLAND (McGarigal and Marks 1995)	Direct thermal is unsuitable habitat, but breed in areas with thermal influence. (Koch and Peterson 1995)
	1988 Fire perimeter	Burn	USGS	+	Fire parameter from 1988	Count of cells	Increased movement post fire (Hossack and Corn 2007)
Topo	Elevation (m)	Elev	SRTM	-	Elevation in meters	Average elevation	Regulates breeding phenology, affecting gene flow (Giordano et al. 2007)
	Topographic roughness[2]	Rough	SRTM	-	Variance in elevation for a given window size	(Riley et al. 1999)	Topographically complex areas may make dispersal energetically expensive
	Elevation relief ratio[2]	Err	SRTM	-	Index of elevational complexity	(Evans 1972)	Topographically complex areas may make dispersal energetically expensive
	Relative slope position[3]	Rsp	SRTM	+	Relative position between valley floor and ridge top	Appendix 7.2	Sites of similar slope position may be more connected
	Hierarchical slope position[3]	Hsp	SRTM	+	Scale decomposition of slope position	Appendix 7.2	Exposed areas may impede dispersal (Bartelt et al. 2004).
	Ridges	Ridge	SRTM	-	Count of ridge cells	Ridges derived from hsp	May limit dispersal (Funk et al. 1999)
Temp-Moist	Heat Load Index	Hli	SRTM	-	Cool – hot values (slope-aspect transformation)	(McCune and Keon 2002	Toad have little physiological control over water loss, hot areas impede dispersal (Bartelt and Peterson 2005)
	Compound topographic index	cti	SRTM	+	Measure of wetness: flow accumulation by catchment size	(Moore et al. 1993)	Wetness may enhance dispersal. (Bartelt and Peterson 2005).
	Hot-dry slopes	Ssina	SRTM	-	Indicates temperature-moisture	(Stage 1976)	Hot-dry slopes areas impede dispersal (Bartelt and Peterson 2005)
	Curvature	Crv	SRTM	+	Curvature in elevation raster	(Zeverbergen and Thorne 1987)	High curvature may retain moisture while low curvature may be exposed.
	Solar insolation	Inso	SRTM	+	Measure of solar energy Breeding: April – July	(Fu and Rich 1999)	Year – measure of site productivity. Breeding season solar insolation relates energy needed for tadpole rearing
	Growing season precipitation	gsp	Spline	+	Average growing season precip	(Rehfeldt 2006)	May affect breeding and dispersal (Bartelt and Peterson 2005)
	Mean annual precip	Map	Spline	+	Average annual precipitation	(Rehfeldt 2006)	Year-round precipitation, including snow pack, may influencing breeding timing and dispersal (Corn 2003)
	Frost free period	Ffp	Spline	+	Julian date of last freeze to date of first freeze	(Rehfeldt 2006)	Later last freeze means a shortened breeding season, which may result in restricted gene flow (Palo et al. 2003).
	Degree days > 5° C	Dd5	Spline	-	Sum of degrees for all days > 5° C	(Rehfeldt 2006)	High temperatures increase desiccation risk (Bartelt and Peterson 2005)

Some metrics were calculated at multiple window sizes [1] Percent of landscape (PLAND) thermal at 0, 13, and 26 cell radii. [2]Calcualted at 3X3, 15X15, and 27X27 cell window sizes. [3]Decomposed from 0 – 27X27 cell window size.

Appendix 7.2

Definitions for newly derived topographic morphology metrics.

We define relative slope position the elevation of a site relative to the valley floor and ridge top by

$$\mathbf{rsp = ((z_s - z_v)/(z_r - z_v)) *100 +0.05}$$

where z_s = elevation at a site, z_v = elevation of the valley floor, and z_r = elevation of the ridge.

Hierarchical slope position is a cross-scale slope decomposition function calculated by applying moving-windows with increasing window sizes to the DEM (3X3 – 27X27 windows). For the smallest window size, hsp calculates the difference between elevation of the center cell (focal cell) and mean elevation. The mean is then standardized and becomes the base grid, iterated to the largest window size (see pseudo-code).

hsp pseudo-code:

Definitions
tmp = original DEM
#Window size range (min to max) and increment (i)#
min = 3
max = 27
i=2
s=min
mean(sx) = global mean of grid sx
stdv(sx) = global standard deviation for grid sx

Begin loop:
xs = tmp – FOCALMEAN
#for each cell, value in the tmp grid minus mean for window#

*scale_s = 100 * (xs – mean(xs)/stdv(sx)*)
s = s + i
#increase window size by one increment#
kill tmp
rename sx temp
kill scale_s
Go to beginning of loop

Literature Cited

Arens, P., T. van der Sluis, W. P. van't Westende, B. Vosman, C. C. Vos, and M. J. M. Smulders. 2007. Genetic population differentiation and connectivity among fragmented Moor frog (*Rana arvalis*) populations in The Netherlands. Landscape Ecology 22:1489-1500.

Bartelt, P., and C. R. Peterson. 2005. Physical modeling operative temperature and evaporative water loss in amphibians. Journal of Thermal Biology 30:93-102.

Bartelt, P., C. R. Peterson, and R. W. Klaver. 2004. Sexual differences in the movements and habitats selected by Western toads in southeastern Idaho. Herpetologica 60:455-467.

Corn, P. S. 2003. Amphibian breeding and climate change: importance of snow in the mountains. Conservation Biology 17:622-625.

Evans, I. S. 1972. General geomorphometry, detivatives of altitude, and descriptive statistics. *in* R. J. Chorley, editor. Spatial Analysis in geomorphology. Harper & Row, New York.

Fu, P., and P. M. Rich. 1999. Design and implementation of the Solar Analyst: an ArcView extension for modeling solar radiation at landscape scales. *in* 19th Annual ESRI User Conference. ESRI, San Diego, CA.

Funk, W. C., D. A. Tallmon, and F. W. Allendorf. 1999. Small effective population size in the long-toed salamander. Molecular Ecology 8:1633-1640.

Giordano, A. R., B. J. Ridenhour, and A. Storfer. 2007. The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactylum*). Molecular Ecology 16:1625-1637.

Hossack, B. R., and P. S. Corn. 2007. Responses of pond-breeding amphibians to wildfire: short-term patterns in occupancy and colonization. Ecological Applications 17:1403-1410.

Koch, E. D., and C. R. Peterson. 1995. Amphibians and Reptiles of Yellowstone and Grand Teton National Parks. University of Utah Press, Salt Lake City.

Manier, M. K., and S. J. Arnold. 2006. Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. Proceedings of the Royal Society B 273:3001-3009.

McGarigal, K., and B. J. Marks. 1995. FRAGSTATS. Spatial analysis program for quantifying landscape structure. USDA Forest Service General Technical Report PNW-GTR-351.

Moore, I., P. Gessler, G. Nielsen, and G. Petersen, editors. 1993. Terrain attributes and estimation methods and scale effects. Wiley, London.

Palo, J. U., R. B. O'Hara, A. T. Laugen, A. Laurila, C. R. Primmer, and J. Merila. 2003. Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. Molecular Ecology 12:1963–1978.

Rehfeldt, G. E. 2006. A spline model of climate for the Western United States. RMRS-GTR-165, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Riley, S. J., S. D. DeGloria, and R. Elliot. 1999. A terrain ruggedness index that quantifies topographic heterogeniety. Intermountain Journal of Sciences 5:23-27.

Stage, A. 1976. An expression of the effects of aspect, slope, and habitat type on tree growth. Forest Sciences 22:457-469.

Zeverbergen, L. W., and C. R. Thorne. 1987. Quantitative analysis of land surface topography. Earth Surface Processes and Landforms 12:47-56.

Appendix 8

Full table of model results (expanded from Table 6). Landscape genetic models of *B. boreas* connectivity in Yellowstone National Park. Hypothesis – overall hypothesis group (Process, Scale, or Hierarchical). Process is the tested ecological processs ("Class"): habitat permeability, topographic morphology, temperature-moisture, or an interaction of all process (All). Scale is the buffer size around each vector connecting sites (meters). Distance is the pair-wise genetic distance measure. We present several validation statistics: R2 - amount of variation explained (a pseudo-R2), rmse - root mean squared error, rmsd - root mean standard deviation, resd.mean - mean of the residuals, resd.stdv - standard deviation of the residuals, resd.skew – skewness of the residuals, reds.kurt 0 kurtosis of the residuals, MSE is mean squared error, Sig - overall model significance. Model is the selected model with metrics in order of importance (abbreviations in Table 5). Number after a variable indicates the scale (in 30 m pixel window size) for multi-scale variables. Number after "_" for cross scale models indicates buffer size. Threshold is the MIR threshold for the model.

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Process	Habitat	FST	33.480	0.03	0.385	8.00E-04	0.0295	0.5562	-0.0027	0.00392	p < 0.001	meadow, canopy, wetland, distance, burn, forest, therm720, imperv	70
		DPS	51.797	0.031	0.289	-0.001	0.0307	0.4117	0.2911	0.00546	p < 0.001	burn, distance, canopy, water, forest, meadow, imperv, wetland, therm360	70
		DAN	44.400	0.036	0.343	-0.0016	0.0359	0.4541	-0.2608	0.00616	p < 0.001	distance, forest, burn, meadow, canopy, water	70
		CAS	50.838	0.031	0.343	-9.00E-04	0.0312	0.2272	-0.3687	0.00408	p < 0.001	burn, distance, forest, canopy	80

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Temp Moist	FST	39.939	0.029	0.378	7.00E-04	0.029	0.4669	0.1759	0.00354	p < 0.001	map, gsp, distance, hli, dd5, ffp	70
		DPS	55.309	0.034	0.322	-4.00E-04	0.0343	0.0464	-0.2954	0.00466	p < 0.001	gsp, distance	70
		DAN	47.705	0.038	0.362	-3.00E-04	0.0379	0.3433	-0.3016	0.00576	p < 0.001	gsp, distance, map	70
		CAS	54.043	0.031	0.347	-3.00E-04	0.0315	0.0523	-0.4234	0.00381	p < 0.001	distance, gsp	80
	Topo	FST	34.357	0.028	0.37	9.00E-04	0.0284	0.5309	-0.0101	0.00389	p < 0.001	elev, distance, rough15, rsp, rough3, rough27, err27, hsp, err15, err3	60
		DPS	53.570	0.031	0.293	-9.00E-04	0.0311	0.5469	0.4277	0.00524	p < 0.001	hsp, elev, rough15, err27, distance,rough27, err3, err15, rough3	70
		DAN	42.678	0.036	0.348	-2.00E-04	0.0364	0.5355	0.0564	0.00631	p < 0.001	rough27, distance, err27, elev, hsp, rough15, rough3, err3, err15, ridge	60
		CAS	51.380	0.028	0.31	-0.001	0.0281	0.2353	-0.1505	0.00405	p < 0.001	hsp, err27, rough27, elev, rough15, distance, err3, err15	70

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	38.420	0.026	0.343	8.00E-04	0.0263	0.5854	0.1614	0.00364	p < 0.001	meadow, wetland, gsp, map, canopy, dd5, elev, rough3, rough15, ffp, burn, errr27, imperv, distance, rsp, err15, hli, rough27, therm720, forest, hsp	50
		DPS	56.284	0.03	0.282	-2.00E-04	0.03	0.3196	-0.0205	0.00501	p < 0.001	gsp, distance, ssina, rough27, rough15, burn, canopy, hsp, err27, dd5, err15, forest, ffp	60
		DAN	45.106	0.033	0.312	7.00E-04	0.0327	0.4946	0.1801	0.00602	p < 0.001	ssina, err27, rough15, distance, burn, rough27, gsp, dd5, ffp, hsp, ridge, forest, meadow, insob, imperv, err15, canopy, rough3, map, err3, elev, rsp, wetland, water, insoy, hli, cti	40
		CAS	53.462	0.026	0.288	-0.001	0.0262	0.2446	-0.116	0.00384	p < 0.001	ssina, burn, err27, distance, rough15, ridge, hsp, canopy, err3, gsp, ffp, dd5, err15, forest, rough27, meadow, imperv	40

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Scale 60	Habitat	FST	33.851	0.029	0.377	6.00E-04	0.0289	0.525	0.1753	0.00387	p < 0.001	meadow, canopy, burn, imperv, distance, forest, wetland, therm720	70
		DPS	51.251	0.03	0.286	-1.00E-04	0.0304	0.4272	0.2653	0.00548	p < 0.001	burn, distance, canopy, water, forest, imperv, meadow	60
		DAN	40.754	0.035	0.339	-6.00E-04	0.0355	0.4781	-0.1316	0.00653	p < 0.001	burn, distance, canopy, water, forest, imperv, meadow	70
		CAS	49.571	0.028	0.305	-6.00E-04	0.0277	0.2331	0.0146	0.00413	p < 0.001	burn, distance, canopy, forest, meadow, imperv	80
	Temp Moist	FST	41.598	0.029	0.384	8.00E-04	0.0294	0.499	0.2338	0.00344	p < 0.001	map, gsp, distance, dd5, ffp	70
		DPS	54.557	0.033	0.309	7.00E-04	0.0329	0.1528	-0.2818	0.00518	p < 0.001	gsp, distance, ssina, map, dd5	50
		DAN	45.120	0.039	0.372	3.00E-04	0.039	0.4393	0.0296	0.00604	p < 0.001	gsp, distance, map, dd5	60
		CAS	53.752	0.029	0.317	4.00E-04	0.0288	0.0834	-0.296	0.00383	p < 0.001	gap, distance, ssina, map	70
	Topo	FST	35.185	0.031	0.404	8.00E-04	0.031	0.4621	0.1286	0.00380	p < 0.001	elev, rough15, rough27, distance, rough3, err15	80
		DPS	53.081	0.033	0.312	1.00E-04	0.0332	0.3206	0.0468	0.00532	p < 0.001	elev, hsp, rough15, distance, err27, rough27, err15	50
		DAN	42.164	0.036	0.348	-3.00E-04	0.0365	0.5326	0.1599	0.00633	p < 0.001	elev, hsp, distance, rough27, rough15, err27, rough3, err15, rsp, ridge	50
		CAS	50.585	0.029	0.325	2.00E-04	0.0295	0.1609	-0.1356	0.00407	p < 0.001	elev, hsp, distance, rough15, err27, rough27	60

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	42.150	0.027	0.347	3.00E-04	0.0266	0.5079	0.3514	0.00339	p < 0.001	meadow, gsp, err15, imperv, dd5, map, canopy, rough15	50
		DPS	56.866	0.03	0.282	3.00E-04	0.0299	0.2757	-0.1683	0.00490	p < 0.001	gsp, ssina, hsp, burn, rough15, err27, canopy, dd5, ffp, distance, rough27	60
		DAN	45.891	0.033	0.315	6.00E-04	0.033	0.4075	-0.2027	0.00597	p < 0.001	hsp, ssina, rough15, err27, distance, gsp, burn, ffp, rough27, meadow, dd5, canopy, rough3, elev, water, forest, insob, err15, wetland, rsp, map, insoy, ridge, hli	50
		CAS	55.204	0.026	0.287	-4.00E-04	0.0261	0.1722	-0.3302	0.00370	p < 0.001	ssina, burn, rough15, canopy, err27, distance, imperv, ffp, water, gsp, meadow, dd5, elev	40
Scale 120	Habitat	FST	34.751	0.029	0.38	8.00E-04	0.0292	0.5236	-4.00E-04	0.00387	p < 0.001	imperv, meadow, canopy, distance, burn, forest, wetland, therm720	70
		DPS	51.993	0.033	0.31	-9.00E-04	0.0329	0.4414	0.2972	0.00541	p < 0.001	burn, distance, imperv, water, canopy, meadow, forest	60
		DAN	44.428	0.035	0.337	-0.0011	0.0353	0.3935	-0.2897	0.00613	p < 0.001	imperv, burn, distance, forest, meadow, water, canopy, wetland	50
		CAS	52.343	0.028	0.306	-9.00E-04	0.0277	0.2109	0.0458	0.00396	p < 0.001	burn, imperv, distance, forest, water, meadow, canopy, wetland	50

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Temp Moist	FST	40.043	0.029	0.377	0.0011	0.0289	0.4858	0.2742	0.00356	p < 0.001	map, gsp, distance, ffp, hli	50
		DPS	53.744	0.034	0.325	0.0011	0.0345	0.1015	-0.3304	0.00529	p < 0.001	gsp, distance, ssina, map	50
		DAN	46.239	0.038	0.36	1.00E-04	0.0377	0.4464	-0.1261	0.00591	p < 0.001	gsp, distance, map, dd5, ffp	60
		CAS	53.314	0.032	0.348	1.00E-04	0.0315	0.1155	-0.5228	0.00386	p < 0.001	gsp, distance	90
	Topo	FST	35.134	0.031	0.407	0.0013	0.0312	0.4682	0.1761	0.00379	p < 0.001	elev, rough15, rough27, distance, rough3, err15	80
		DPS	50.980	0.033	0.311	-7.00E-04	0.0331	0.334	0.0201	0.00557	p < 0.001	elev, hsp, rough15, err27, distance, rough27, ridge, rough3, err15	40
		DAN	43.646	0.037	0.357	0.0012	0.0374	0.431	-0.2178	0.00617	p < 0.001	elev, hsp, distance, rough27, rough15, err27, ridge, rough3, rsp, err15	60
		CAS	49.957	0.03	0.332	-1.00E-04	0.0302	0.1888	-0.3043	0.00413	p < 0.001	elev, hsp, distance, err27, rough15	70

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	40.689	0.027	0.347	3.00E-04	0.0266	0.5579	0.1853	0.00351	p < 0.001	meadow, imperv, gsp, err15, canopy, map, dd5, rough3, rough27	50
		DPS	56.914	0.03	0.287	-4.00E-04	0.0305	0.2035	-0.1915	0.00487	p < 0.001	gsp, ssina, hsp, rough15, burn, err27, canopy, dd5, ffp, distance, imperv	60
		DAN	45.854	0.033	0.315	1.00E-04	0.033	0.4298	-0.2547	0.00600	p < 0.001	ssina, hsp, err27, imperv, rough15, distance, burn, ffp, ridge, meadow, dd5, rough27, gsp, water, canopy, insob, rsp, rough3, forest, elev, wetland, map, err15	50
		CAS	54.784	0.027	0.296	4.00E-04	0.0269	0.2306	-0.3023	0.00372	p < 0.001	ssina, hsp, burn, err27, rough15, imperv, canopy, distance, ffp	60
Scale 240	Habitat	FST	31.496	0.031	0.406	0.0013	0.0311	0.5473	0.25	0.00400	p < 0.001	imperv, meadow, canopy, forest, distance, wetland, burn, therm720, water, therm360	40
		DPS	49.502	0.033	0.306	-3.00E-04	0.0326	0.3678	0.4328	0.00570	p < 0.001	distance, burn, meadow, forest, canopy, water, imperv	50
		DAN	40.702	0.037	0.354	-4.00E-04	0.0371	0.3977	-0.1671	0.00649	p < 0.001	distance, imperv, forest, meadow, water, burn, canopy	60
		CAS	48.176	0.029	0.319	-5.00E-04	0.029	0.1604	-0.0374	0.00425	p < 0.001	distance, burn, imperv, water, forest, meadow, canopy	60

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Temp Moist	FST	40.199	0.029	0.378	7.00E-04	0.029	0.5404	0.2601	0.00352	p < 0.001	map, gsp, distance, ffp, dd5, insob	60
		DPS	56.558	0.034	0.324	8.00E-04	0.0345	0.1667	-0.5071	0.00498	p < 0.001	ssina, gsp, distance, map	60
		DAN	47.418	0.039	0.373	0.001	0.0391	0.4291	-0.3788	0.00573	p < 0.001	ssina, gsp, distance	80
		CAS	56.654	0.03	0.33	2.00E-04	0.03	0.0642	-0.5057	0.00355	p < 0.001	ssina, distance, gsp	70
	Topo	FST	32.773	0.03	0.394	0.001	0.0302	0.5339	0.0496	0.00398	p < 0.001	elev, ridge, distance, rough3, rough15, rough27, err27, rsp, err3, err15, hsp	10
		DPS	51.655	0.032	0.301	-9.00E-04	0.032	0.4302	0.2721	0.00557	p < 0.001	hsp, elev, err27, err3, ridge, distance, rough27, rough15, rough3, err15	60
		DAN	44.368	0.038	0.36	-4.00E-04	0.0377	0.4697	0.0726	0.00612	p < 0.001	rough27, distance, hsp, ridge, rough15, err27	90
		CAS	51.139	0.029	0.322	-8.00E-04	0.0292	0.2034	0.0079	0.00402	p < 0.001	hsp, rough27, ridge, elev, rough15, err3, err27	90

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	38.043	0.028	0.37	7.00E-04	0.0284	0.6517	0.3501	0.00366	p < 0.001	meadow, imperv, ridge, gsp, map, wetland, dd5, canopy, rough3, rough15, burn, ffp, elev, rough27, rsp	50
		DPS	58.325	0.031	0.295	-5.00E-04	0.0314	0.1621	-0.1929	0.00477	p < 0.001	ssina, gsp, ridge, canopy, imperv, burn	50
		DAN	46.959	0.036	0.34	-4.00E-04	0.0356	0.4581	-0.1957	0.00584	p < 0.001	ssina, ridge, imperv, err27, gsp, rough15, hsp, distance	60
		CAS	53.297	0.029	0.315	-5.00E-04	0.0286	0.2086	-0.2964	0.00387	p < 0.001	ssina, ridge, burn, err27, hsp, imperv, rough15, rough3	50
Scale 480	Habitat	FST	36.181	0.03	0.391	2.00E-04	0.03	0.3654	0.1794	0.00372	p < 0.001	imperv, meadow, wetland	70
		DPS	50.139	0.034	0.324	-3.00E-04	0.0344	0.3336	0.3994	0.00565	p < 0.001	distance, water, imperv, meadow, burn, forest, canopy, wetland	70
		DAN	42.361	0.041	0.39	-6.00E-04	0.0409	0.407	-0.2075	0.00633	p < 0.001	distance, water, imperv, meadow, burn, forest, canopy, wetland	90
		CAS	49.014	0.031	0.339	-8.00E-04	0.0307	0.2038	-0.0411	0.00421	p < 0.001	meadow, distance, imperv	70
	Temp Moist	FST	39.721	0.031	0.409	9.00E-04	0.0313	0.4904	0.1838	0.00354	p < 0.001	map, gsp, distance, ffp	80
		DPS	57.762	0.034	0.321	-4.00E-04	0.0341	0.1104	-0.5304	0.00479	p < 0.001	ssina, gsp, distance	60
		DAN	48.522	0.038	0.366	-4.00E-04	0.0383	0.3629	-0.3901	0.00569	p < 0.001	ssina, gsp, distance	80
		CAS	56.068	0.03	0.332	-4.00E-04	0.0301	0.0824	-0.5606	0.00362	p < 0.001	ssina, distance, gsp	60

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Topo	FST	32.582	0.032	0.419	6.00E-04	0.0322	0.4517	0.0909	0.00394	p < 0.001	elev, rough3, distance, ridge, rough27, rough15, rsp	80
		DPS	48.892	0.035	0.326	-3.00E-04	0.0347	0.3477	0.0504	0.00578	p < 0.001	elev, err3, hsp, rough27, ridge, distance, err27, rough3	80
		DAN	41.446	0.039	0.37	1.00E-04	0.0387	0.4708	-0.1316	0.00643	p < 0.001	ridge, rough27, distance, rough15, elev, hsp, err27, err3, rough3, err15, rsp	40
		CAS	50.723	0.029	0.323	-8.00E-04	0.0293	0.2274	-0.0666	0.00409	p < 0.001	ridge, elev, hsp, rough15, distance, err27, rough27, err3	90
	Full	FST	40.101	0.028	0.365	8.00E-04	0.028	0.4869	0.1872	0.00354	p < 0.001	meadow, imperv, ridge, gsp, wetland, map, rough3	60
		DPS	56.120	0.035	0.334	-3.00E-04	0.0355	0.1081	-0.3892	0.00502	p < 0.001	ssina, gsp, ridge	60
		DAN	45.088	0.036	0.346	-3.00E-04	0.0363	0.462	-0.0849	0.00604	p < 0.001	ssina, ridge, meadow, gsp, distance, imperv, rough15, burn, rough3, err3, rough27, ffp, err27	50
		CAS	53.496	0.029	0.323	-6.00E-04	0.0293	0.1117	-0.3678	0.00385	p < 0.001	ssina, ridge, burn, imperv, gsp, err27, meadow, distance, rough3, err3, hsp	40

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Scale 960	Habitat	FST	35.903	0.029	0.381	0.0014	0.0292	0.4777	0.0688	0.00380	p < 0.001	wetland, imperv, meadow, canopy, burn, distance, therm 720, therm360, water, forest	50
		DPS	51.380	0.033	0.309	-7.00E-04	0.0328	0.4729	0.6727	0.00553	p < 0.001	water, meadow, distance, burn, forest, imperv, wetland	70
		DAN	43.050	0.038	0.362	-1.00E-04	0.0379	0.4701	-0.1543	0.00627	p < 0.001	water, meadow, distance, burn, wetland, imperv	70
		CAS	51.176	0.029	0.325	-4.00E-04	0.0295	0.2575	-0.1028	0.00402	p < 0.001	water, meadow, distance, burn, forest, wetland, imperv	70
	Temp Moist	FST	40.876	0.03	0.389	8.00E-04	0.0298	0.5603	0.2117	0.00350	p < 0.001	map, gsp, distance, dd5	70
		DPS	57.975	0.034	0.316	-1.00E-04	0.0337	0.1207	-0.196	0.00474	p < 0.001	gsp, ssina, distance	60
		DAN	45.800	0.038	0.367	0	0.0385	0.4556	-0.0994	0.00592	p < 0.001	gsp, distance, ssina	90
		CAS	56.301	0.03	0.327	-1.00E-04	0.0297	0.0776	-0.3122	0.00362	p < 0.001	sinna, distance, gsp	70
	Topo	FST	34.247	0.034	0.441	4.00E-04	0.0339	0.3402	0.1626	0.00388	p < 0.001	ridge, elev, distance	90
		DPS	51.462	0.032	0.304	3.00E-04	0.0324	0.2758	0.125	0.00555	p < 0.001	ridge, distance, elev, rough15, hsp, rough27, err3, err27, rough3, err15, rsp	30
		DAN	42.436	0.043	0.415	-3.00E-04	0.0434	0.4658	-0.2242	0.00629	p < 0.001	ridge, distance, err27	90
		CAS	51.720	0.033	0.359	-3.00E-04	0.0326	0.2554	0.1632	0.00397	p < 0.001	ridge, distance, err27, elev	80

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	42.140	0.028	0.365	8.00E-04	0.028	0.5268	0.3009	0.00343	p < 0.001	meadow, imperv, ridge, wetland, canopy, gsp, map, dd5	50
		DPS	59.988	0.034	0.319	2.00E-04	0.0339	0.1464	-0.1307	0.00454	p < 0.001	ssina, ridge, gsp	60
		DAN	48.804	0.035	0.331	-4.00E-04	0.0347	0.5575	0.1345	0.00563	p < 0.001	ridge, ssina, meadow, scosa, imperv, gsp, rough3, dd5, err27, water, burn, cti, distance	50
		CAS	57.473	0.026	0.29	-2.00E-04	0.0263	0.2186	-0.0752	0.00353	p < 0.001	ssina, ridge, burn, scosa, gsp, meadow, err27, imperv, rough3, distance, water, dd5	40
Cross-scale	Habitat	FST	40.622	0.025	0.324	0.001	0.0249	0.3894	0.2451	0.00352	p < 0.001	imperv_480, wetland_960, imperv_240, meadow_480, meadow_960, distance, imperv_960, canopy, meadow_240, forest, wetland, meadow_60, imperv_120	80
		DPS	52.174	0.033	0.307	-0.0018	0.0326	0.4547	0.3208	0.00541	p < 0.001	distance, burn, burn_60, imperv_240, canopy_60, imperv_120	60
	Temp Moist	FST	39.117	0.03	0.39	0.001	0.0299	0.5309	0.3356	0.00360	p < 0.001	gsp, map_60, map, distance, dd5, ffp	80
		DPS	57.490	0.033	0.314	1.00E-04	0.0334	0.1273	-0.5501	0.00483	p < 0.001	gsp, distance, ssina_480	50
	Topo	FST	36.967	0.032	0.418	5.00E-04	0.032	0.4526	-0.0088	0.00372	p < 0.001	elev, err15_60, distance, ridge_960	80
		DPS	52.515	0.034	0.32	8.00E-04	0.034	0.3088	0.0046	0.00540	p < 0.001	hsp_60, hsp_120, ridge_960, rough27, distance, err27, err27_60	80

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	43.925	0.025	0.327	8.00E-04	0.0251	0.5993	0.1912	0.00334	p < 0.001	meadow_60, meadow_480, imperv_480, meadow_240, ridge_960, canopy, meadow_960, imperv_240, err15_60, ridge_240, dd5_960, err15_120, wetland_960, gsp_60, dd5_60, imperv_120, map_120, imperv_60, canopy_120, imperv_960, gsp_120, gsp_480	70
		DPS	59.087	0.031	0.29	0.0012	0.0308	0.2978	-0.1841	0.00467	p < 0.001	ridge_960, gsp, ssina_480, canopy_60, ssina_240, ssina_960, imperv_240, err27_60	70
Within 0	Habitat	FST	31.723	0.045	0.517	-0.0011	0.0451	0.252	-0.1266	0.00507	p < 0.001	canopy, imperv, therm	40
		DPS	56.699	0.043	0.332	-0.001	0.043	0.2868	0.4798	0.00707	p < 0.001	burn, imperv, meadow, canopy	70
		DAN	50.993	0.038	0.323	-3.00E-04	0.0383	0.4419	0.326	0.00669	p < 0.001	meadow, distance, burn, imperv, forest, canopy, wetland, them	40
		CAS	49.477	0.036	0.337	-8.00E-04	0.0366	0.2686	0.3427	0.00582	p < 0.001	meadow, burn, distance, forest, imperv, canopy, wetland	40
	Temp Moist	FST	25.852	0.043	0.492	0.0027	0.0429	-0.1003	-0.6608	0.00544	p < 0.001	map, distance	60
		DPS	62.484	0.033	0.261	5.00E-04	0.0338	0.0332	-0.1315	0.00620	p < 0.001	gsp, distance, cti, map, ssina, ffp, dd5	30
		DAN	52.750	0.04	0.341	1.00E-04	0.0404	0.0201	0.1615	0.00648	p < 0.001	distance, gsp, map	60
		CAS	53.050	0.036	0.334	-5.00E-04	0.0364	-0.0492	-0.0845	0.00538	p < 0.001	distance, gsp, map	70

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Topo	FST	21.118	0.045	0.522	5.00E-04	0.0456	0.0128	-0.9327	0.00579	p < 0.001	err27, distance, rsp, err15, rough3	40
		DPS	54.845	0.038	0.3	-0.0026	0.0388	0.5793	1.102	0.00741	p < 0.001	err27, elev, hsp, rough27, ridge, rough15, err15, rough3	70
		DAN	47.637	0.041	0.347	-7.00E-04	0.0412	0.4407	0.2892	0.00708	p < 0.001	distance, err3, err27, rough27, rough3, err15, rough15, rsp, hsp	50
		CAS	53.685	0.034	0.312	-0.0012	0.034	0.0657	0.3284	0.00526	p < 0.001	err27, ridge, rough27, err15, hsp, err3, distance, rough15, rough3, rsp	60
	Full	FST	28.640	0.044	0.511	9.00E-04	0.0446	0.0991	-0.663	0.00526	p < 0.001	canopy, err27	80
		DPS	62.052	0.033	0.257	-0.0021	0.0333	0.1434	0.7708	0.00615	p < 0.001	gsp, err27, meadow, burn, hsp, imperv, canopy, err15	40
		DAN	49.372	0.038	0.324	-8.00E-04	0.0384	0.305	0.1744	0.00692	p < 0.001	meadow, burn, gsp, err27, forest, distance, ridge, err3, imperv, err15, rough27, rough15, wetland, rough3, ssina, hsp, rsp	40
		CAS	53.490	0.03	0.282	-3.00E-04	0.0307	0.2524	0.9332	0.00530	p < 0.001	meadow, burn, ridge, gsp, err27, forest, distance, hsp, err3, rough27, ssina, wetland, canopy, rough15, err15, imperv, rough3, cti, rsp, dd5	30

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Within 60	Habitat	FST	26.918	0.042	0.485	-2.00E-04	0.0424	0.105	-0.8016	0.00543	p < 0.001	canopy, imperv, distance, burn	60
		DPS	52.530	0.05	0.391	-0.0036	0.0506	0.3762	0.4017	0.00768	p < 0.001	burn, meadow, imperv	80
		DAN	46.107	0.039	0.335	-0.0018	0.0397	0.4834	0.3261	0.00750	p < 0.001	burn, meadow, distance, imperv, forest, canopy, wetland, therm	40
		CAS	51.721	0.038	0.355	-0.0018	0.0386	0.3072	-0.1246	0.00558	p < 0.001	burn, meadow, canopy	60
	Temp Moist	FST	31.173	0.038	0.44	0.0031	0.0383	-0.0864	-0.4905	0.00506	p < 0.001	map, distance	70
		DPS	65.567	0.031	0.239	0.0018	0.0309	0.0754	0.0038	0.00554	p < 0.001	gsp, distance, map, ssina, dd5, cti	30
		DAN	55.527	0.037	0.315	0.001	0.0374	0.1466	0.0416	0.00608	p < 0.001	gsp, distance, map	70
		CAS	55.124	0.031	0.291	5.00E-04	0.0317	-0.0269	-0.4353	0.00531	p < 0.001	gsp, cti, distance, map, ssina	60
	Topo	FST	30.369	0.041	0.475	1.60E-03	0.0415	-0.0933	-0.8186	0.00510	p < 0.001	err27, distance	70
		DPS	53.536	0.042	0.325	-0.0019	0.0421	0.3392	0.4142	0.00753	p < 0.001	err27, rough15, rough3, rough27, elev	80
		DAN	45.027	0.04	0.339	-2.50E-03	0.0401	0.2751	-0.2747	0.00757	p < 0.001	err27, hsp, distance, rough27, ridge, rough15, err15, rough3	80
		CAS	46.921	0.037	0.345	-0.0026	0.0375	0.4211	0.924	0.00613	p < 0.001	err27, distance, ridge, rough27, rough3, hsp, elev, rough15, err15	50

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	28.851	0.034	0.392	1.20E-03	0.0342	0.1467	-0.6561	0.00520	p < 0.001	sinna, ffp, wetland, imperv, hli, rough15, rough3, burn, map, err15, distance, elev, dd5, canopy, err27, meadow	30
		DPS	62.204	0.03	0.234	-0.0017	0.0303	0.1782	0.4862	0.00627	p < 0.001	distance, rough27, canopy, forest, dd5, err27, err15, ssina, burn, imperv, meadow	30
		DAN	51.536	0.034	0.287	-4.00E-04	0.034	0.4157	0.2179	0.00657	p < 0.001	gsp, meadow, distance, err27, burn, hsp, imperv, forest, canopy, rough27, rough15	50
		CAS	58.182	0.031	0.29	-5.00E-04	0.0316	0.3978	0.9553	0.00483	p < 0.001	meadow, gsp, burn, err27, ssina, forest, canopy	60
Within 120	Habitat	FST	24.062	0.043	0.499	9.00E-04	0.0436	0.1832	-0.7421	0.00563	p < 0.001	imperv, canopy, burn, distance	60
		DPS	53.125	0.044	0.346	-0.003	0.0447	0.4771	0.3662	0.00754	p < 0.001	burn, meadow, imperv, canopy	70
		DAN	46.837	0.041	0.35	1.20E-03	0.0415	0.5414	0.4057	0.00719	p < 0.001	burn, meadow, forest, distance, imperv, therm, wetland, canopy	40
		CAS	49.597	0.036	0.33	-0.0017	0.0359	0.3192	0.0797	0.00572	p < 0.001	burn, meadow, therm, forest, distance, imperv, wetland, canopy	30

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Temp Moist	FST	29.206	0.038	0.442	0.004	0.0384	-0.1515	-0.4829	0.00521	p < 0.001	distance, map	70
		DPS	65.732	0.03	0.231	5.00E-04	0.03	0.1662	-0.0635	0.00585	p < 0.001	gsp, distance, map, cti, dd5, ssina, ffp	20
		DAN	54.481	0.035	0.297	5.00E-04	0.0353	0.098	-0.4375	0.00630	p < 0.001	distance, gsp, map	50
		CAS	58.147	0.032	0.293	0.001	0.0319	0.0186	-0.4485	0.00477	p < 0.001	gsp, distance, ssina, map, cti	50
	Topo	FST	26.724	0.044	0.514	3.00E-04	0.0449	-0.0641	-0.8182	0.00541	p < 0.001	err27, distance	80
		DPS	57.983	0.033	0.261	-0.0034	0.0336	0.2887	0.78	0.00677	p < 0.001	elev, err27, hsp, rough27, rough15, ridge	80
		DAN	47.123	0.049	0.42	-0.0016	0.0498	0.4856	0.2309	0.00724	p < 0.001	distance, ridge, err27	90
		CAS	49.763	0.035	0.321	-0.0023	0.0349	0.2978	0.4283	0.00577	p < 0.001	err27, hsp, distance, ridge, rough27, rough15	70
	Full	FST	28.605	0.038	0.435	1.90E-03	0.038	0.1915	-0.8132	0.00528	p < 0.001	imperv, dd5, map, meadow, err15, err27, canopy	40
		DPS	64.864	0.03	0.231	-0.0025	0.0299	0.1034	0.0703	0.00582	p < 0.001	err15, forest, err27, dd5, hsp, imperv, ffp, burn, meadow, gsp	30
		DAN	52.332	0.032	0.276	-1.20E-03	0.0327	0.267	0.1575	0.00638	p < 0.001	gsp, meadow, burn, forest, hsp, distance, err27, ridge, imperv, map, wetland, rough15, ssina, canopy, dd5, ffp, rough27, err15, cti, rough3	30
		CAS	56.208	0.038	0.35	-8.00E-04	0.0381	-0.0242	0.4275	0.00502	p < 0.001	burn, gsp, meadow, err27	70

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Within 240	Habitat	FST	26.055	0.045	0.532	6.00E-04	0.0456	0.0207	-0.6683	0.00523	p < 0.001	canopy, burn	80
		DPS	49.241	0.045	0.357	-0.0021	0.0455	0.153	0.1068	0.00786	p < 0.001	burn, meadow, imperv, distance, barren, therm, water	40
		DAN	47.537	0.044	0.382	-1.90E-03	0.0446	0.0406	-0.8437	0.00690	p < 0.001	burn, meadow, imperv, therm	50
		CAS	48.937	0.04	0.376	-2.70E-03	0.0403	-0.2156	-0.556	0.00573	p < 0.001	burn, meadow, imperv, therm	40
	Temp Moist	FST	27.494	0.042	0.492	0.0025	0.0421	-0.0395	-0.6641	0.00514	p < 0.001	map, distance	90
		DPS	57.862	0.041	0.322	-4.00E-04	0.041	-0.1032	-0.5992	0.00660	p < 0.001	distance, ssina, gsp	50
		DAN	45.012	0.043	0.37	1.40E-03	0.0432	0.1161	-0.5853	0.00716	p < 0.001	gsp, distance, hli, ssina, map, cti	50
		CAS	52.799	0.035	0.329	-7.00E-04	0.0353	-0.1131	-0.6668	0.00531	p < 0.001	ssina, gsp, distance, map	60
	Topo	FST	20.644	0.048	0.57	4.00E-04	0.0488	-0.1364	-0.6375	0.00564	p < 0.001	err27, distance	80
		DPS	55.013	0.038	0.298	-0.0011	0.0379	0.0496	0.2623	0.00714	p < 0.001	err27, elev, rough15, err3, hsp	80
		DAN	38.768	0.048	0.416	-0.002	0.0486	0.1066	-0.2511	0.00807	p < 0.001	err3, ridge, distance, err27	80
		CAS	45.994	0.038	0.357	5.00E-04	0.0383	-0.0472	0.0519	0.00595	p < 0.001	err27, ridge, rough27, err3, rough15, rsp, distance, rough3, hsp	70
	Full	FST	25.765	0.043	0.501	7.00E-04	0.0429	-0.0274	-0.7864	0.00529	p < 0.001	meadow, map, imperv, canopy, err27	70
		DPS	59.429	0.034	0.271	-7.00E-04	0.0346	0.4554	0.3136	0.00633	p < 0.001	rsp, ffp, imperv, err15, ssina, rough15, err27, dd5, hsp, err3, burn, meadow, gsp	40
		DAN	44.846	0.04	0.348	0	0.0407	0.3529	-0.1642	0.00739	p < 0.001	gsp, burn, meadow, distance, err27	60
		CAS	53.496	0.04	0.375	-0.0015	0.0402	0.0019	-0.4845	0.00521	p < 0.001	burn, meadow, ssina, err27, rsp	40

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Within 480	Habitat	FST	37.814	0.035	0.4	0.002	0.0348	-0.084	-0.7897	0.00462	p < 0.001	imperv, canopy	60
		DPS	59.335	0.034	0.266	4.00E-04	0.0345	0.3656	1.0157	0.00662	p < 0.001	imperv, burn, meadow, water, wetland, distance	40
		DAN	61.403	0.035	0.298	-2.00E-04	0.0353	0.0268	-0.3478	0.00528	p < 0.001	meadow, imperv	90
		CAS	54.678	0.036	0.33	0	0.0359	-0.0935	-0.3171	0.00514	p < 0.001	meadow, imperv	90
	Temp Moist	FST	30.273	0.04	0.458	0.0013	0.04	-0.1107	-0.6138	0.00513	p < 0.001	map, distance	60
		DPS	61.492	0.033	0.258	5.00E-04	0.0334	-0.0953	-0.3595	0.00622	p < 0.001	ffp, map, dd5, hli, ssina, distance, cti, gsp	30
		DAN	50.718	0.038	0.327	0.0011	0.0387	0.0367	-0.2725	0.00667	p < 0.001	distance, gsp, cti, hli, map, ssina, ffp, dd5	20
		CAS	51.584	0.035	0.324	7.00E-04	0.0352	-0.0576	-0.3126	0.00551	p < 0.001	gsp, distance, hli, ssina	70
	Topo	FST	20.189	0.047	0.543	0.0013	0.0474	0.0044	-0.8439	0.00585	p < 0.001	elev, distance, rough3, err15, rsp, err27	50
		DPS	51.163	0.042	0.327	-0.0016	0.0423	0.2663	0.541	0.00795	p < 0.001	elev, rough15, err3, hsp, err27	80
		DAN	42.630	0.042	0.361	-0.0015	0.0428	0.1334	-0.3827	0.00783	p < 0.001	err3, err27, err15, hsp, distance, ridge, rough27, rough15	80
		CAS	46.671	0.042	0.394	-0.0018	0.0428	-0.1042	-0.0433	0.00615	p < 0.001	rough, ridge, err3, rough15, err27	80
	Full	FST	42.986	0.034	0.39	2.30E-03	0.034	-0.0285	-0.4731	0.00420	p < 0.001	err27, burn	80
		DPS	72.508	0.032	0.246	-0.0012	0.0319	-9.00E-04	0.4466	0.00450	p < 0.001	gsp, meadow, imperv	60
		DAN	63.342	0.034	0.292	-1.00E-04	0.0346	0.0089	-0.3821	0.00499	p < 0.001	meadow, imperv	60
		CAS	55.500	0.035	0.322	-2.50E-03	0.0349	-0.0495	-0.01	0.00506	p < 0.001	burn, meadow, imperv, ssina	60

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Within 960	Habitat	FST	30.146	0.042	0.49	0.0027	0.0427	0.0523	-1.0051	0.00513	p < 0.001	imperv, meadow	60
		DPS	65.280	0.036	0.283	-0.0033	0.0365	0.6712	1.5063	0.00571	p < 0.001	distance, meadow, wetland, water, burn, forest, imperv	50
		DAN	60.962	0.035	0.294	-2.10E-03	0.0349	0.4714	0.2547	0.00536	p < 0.001	distance, meadow, water, forest, burn, wetland, imperv	40
		CAS	58.663	0.032	0.299	-2.20E-03	0.0325	0.3975	-0.1924	0.00478	p < 0.001	distance, water, wetland, burn, forest, meadow, imperv	70
	Temp Moist	FST	31.706	0.033	0.379	0.0036	0.0329	0.0887	-0.4737	0.00505	p < 0.001	distance, map, dd5	60
		DPS	63.590	0.035	0.276	5.00E-04	0.0358	-0.3252	-0.4376	0.00583	p < 0.001	gsp, cti, ssina	50
		DAN	51.089	0.036	0.307	-1.00E-04	0.0365	0.1769	0.0271	0.00679	p < 0.001	distance, gsp, ssina, cti, map, hli, dd5	40
		CAS	53.408	0.035	0.327	-6.00E-04	0.0356	0.0088	-0.0706	0.00530	p < 0.001	distance, gsp, ssina	70
	Topo	FST	22.204	0.042	0.482	1.00E-03	0.0421	0.0192	-0.745	0.00575	p < 0.001	rough27, rsp, elev, rough3, err15, ridge, distance, err27	30
		DPS	55.175	0.04	0.312	-7.00E-04	0.0404	0.8133	1.8855	0.00739	p < 0.001	rough3, err15, err27, elev, ridge, hsp	90
		DAN	46.629	0.048	0.406	1.00E-04	0.0482	0.3579	-0.0143	0.00722	p < 0.001	distance, ridge, err27	90
		CAS	53.952	0.037	0.341	0.0024	0.0371	0.4414	0.5974	0.00529	p < 0.001	ridge, err27, rsp	70

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	33.581	0.035	0.404	2.80E-03	0.0352	0.1893	-0.5931	0.00493	p < 0.001	burn, elev, err27, dd5, meadow, imperv	30
		DPS	67.649	0.037	0.289	-3.00E-04	0.0375	-0.201	0.956	0.00527	p < 0.001	gsp, imperv	60
		DAN	51.768	0.037	0.316	1.60E-03	0.0374	0.3921	0.4177	0.00654	p < 0.001	ssina, meadow, ridge, gsp, err27, distance, burn, water, imperv, wetland, hsp, err15, forest	50
		CAS	52.257	0.032	0.296	-5.00E-04	0.0323	0.4769	1.3613	0.00545	p < 0.001	meadow, ssina, burn, err27, ridge, forest, gsp, water, err15, rsp, distance, imperv, rough3, err3, wetland, hsp, scosa	40
Within X	Habitat	FST	34.311	0.039	0.452	0.0025	0.0389	0.1844	-0.7824	0.00470	p < 0.001	imperv_960, canopy	
		DPS	63.966	0.036	0.297	-0.0023	0.0367	0.3201	0.023	0.00531	p < 0.001	meadow_480, imperv_480	80
	Temp Moist	FST	22.882	0.039	0.458	2.60E-03	0.0394	0.0705	-0.6492	0.00553	p < 0.001	map_60, distance, hli_60	60
		DPS	62.317	0.033	0.268	-5.00E-04	0.0333	-0.3417	-0.4028	0.00554	p < 0.001	gsp, distance, ssina_60, gsp_240	40
	Topo	FST	30.157	0.045	0.533	2.00E-04	0.0459	0.0583	-0.6993	0.00495	p < 0.001	err27, err27_480, err27_60	80
		DPS	53.003	0.042	0.345	-0.0019	0.0428	0.5229	0.5828	0.00712	p < 0.001	err27, err27_60, rough27_60	70
	Full	FST	38.902	0.037	0.432	0.0013	0.0372	0.1721	-0.6027	0.00438	p < 0.001	imperv_960, err27, canopy	50
		DPS	74.027	0.028	0.228	-0.0011	0.0283	-0.3435	0.619	0.00385	p < 0.001	meadow_480, gsp_960, imperv_480, gsp_120, gsp_480, gsp	60

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Between 0	Habitat	FST	35.075	0.026	0.364	2.00E-04	0.026	0.7348	0.0028	0.00330	p < 0.001	meadow, canopy, wetland, forest, distance, burn, imperv, therm720	60
		DPS	43.484	0.035	0.377	-0.0024	0.0353	0.3017	-0.5366	0.00496	p < 0.001	distance, burn, forest, canopy, water	60
		DAN	33.383	0.035	0.363	-0.0024	0.035	0.5618	-0.3414	0.00621	p < 0.001	distance, forest, burn, canopy, water, meadow, imperv	70
		CAS	40.323	0.028	0.357	-0.0026	0.0281	0.3293	-0.4417	0.00376	p < 0.001	burn, distance, forest, canopy, water, meadow	60
	Temp Moist	FST	41.925	0.027	0.381	0	0.0271	0.8275	0.4689	0.00294	p < 0.001	map, distance, gsp, dd5	70
		DPS	44.578	0.035	0.376	-8.00E-04	0.0353	0.3811	-0.2806	0.00486	p < 0.001	distance, map, gsp, dd5	70
		DAN	32.155	0.043	0.447	-0.0016	0.0432	0.6285	-0.1585	0.00632	p < 0.001	map, distance, dd5, ffp, gsp	70
		CAS	40.333	0.032	0.406	-0.0013	0.0321	0.3964	-0.3401	0.00370	p < 0.001	distance, map, dd5, ffp, gsp	70
	Topo	FST	35.322	0.031	0.442	-8.00E-04	0.0315	0.7486	0.2968	0.00330	p < 0.001	distance, elev	40
		DPS	39.471	0.038	0.402	-0.0015	0.0377	0.5212	0.0576	0.00530	p < 0.001	rough27, distance, err27	70
		DAN	31.239	0.042	0.436	-0.0011	0.0422	0.6622	0.2352	0.00637	p < 0.001	distance, rough27, elev, hsp	50
		CAS	36.658	0.034	0.427	-6.00E-04	0.0338	0.4666	-0.267	0.00395	p < 0.001	distance, rough27, elev, err27	60

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	39.815	0.024	0.335	1.00E-04	0.0239	0.7496	0.3983	0.00306	p < 0.001	meadow, canopy, map, wetland, forest, gsp, dd5, ridge, elev, distance, imperv, ffp, burn, rough27, rough3, rough15	50
		DPS	44.242	0.034	0.363	-0.0018	0.034	0.4897	-0.0727	0.00484	p < 0.001	distance, map, burn, rough27, gsp, err27, forest	70
		DAN	32.835	0.041	0.425	-0.0011	0.0411	0.7391	0.0905	0.00622	p < 0.001	rough27, distance, map, dd5, elev, forest, err27, ffp, burn, imperv, gsp	70
		CAS	38.581	0.028	0.35	-8.00E-04	0.0277	0.4873	-0.2532	0.00384	p < 0.001	burn, distance, ssina, err27, rough27, map, dd5, elev, forest, ridge, canopy, imperv, hsp, cti, ffp, gsp, water, err15	50
Between 60	Habitat	FST	37.388	0.026	0.364	-7.00E-04	0.0259	0.6978	0.0371	0.00320	p < 0.001	meadow, canopy, burn, distance, forest, imperv	70
		DPS	45.240	0.03	0.325	-0.0012	0.0305	0.3821	-0.1607	0.00481	p < 0.001	distance, burn, water, canopy, meadow, imperv, forest	70
		DAN	33.817	0.037	0.38	-0.001	0.0367	0.5711	-0.2427	0.00612	p < 0.001	distance, imperv, forest, burn, water, meadow, canopy	70
		CAS	41.732	0.027	0.346	-0.0015	0.0273	0.3931	-0.2801	0.00363	p < 0.001	distance, burn, imperv, forest, meadow, water, canopy	50

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Temp Moist	FST	42.176	0.027	0.376	3.00E-04	0.0268	0.7053	0.3799	0.00292	p < 0.001	ffp, dd5, distance, gsp, map	70
		DPS	44.544	0.036	0.381	-8.00E-04	0.0357	0.5123	-0.1985	0.00487	p < 0.001	dd5, distance, gsp, map	80
		DAN	32.992	0.042	0.44	-8.00E-04	0.0426	0.6502	-0.0336	0.00625	p < 0.001	dd5, distance, gsp, ffp, map	80
		CAS	42.520	0.032	0.402	-0.0011	0.0318	0.357	-0.3641	0.00357	p < 0.001	dd5, distance, gsp, map	80
	Topo	FST	36.058	0.027	0.378	1.00E-04	0.027	0.7888	0.4194	0.00323	p < 0.001	elev, distance, err15, rough15, rough27, rough3	60
		DPS	39.191	0.037	0.398	-4.00E-04	0.0373	0.445	-0.2411	0.00532	p < 0.001	distance, rough27, elev, err27	80
		DAN	28.234	0.046	0.475	2.00E-04	0.0459	0.6566	-0.0646	0.00671	p < 0.001	distance, rough27, elev, err27	80
		CAS	37.276	0.032	0.409	-6.00E-04	0.0324	0.3822	-0.471	0.00391	p < 0.001	elev, rough27, distance, err27	70
	Full	FST	44.471	0.021	0.294	0	0.0209	0.5856	0.5525	0.00282	p < 0.001	meadow, gsp, err15, imperv, forest, canopy, rough27, ffp, map, dd5, cti, elev, wetland	40
		DPS	44.658	0.037	0.398	-0.0011	0.0373	0.4163	-0.0683	0.00485	p < 0.001	gsp, distance	90
		DAN	33.500	0.039	0.408	-0.0012	0.0395	0.6566	0.0329	0.00620	p < 0.001	imperv, rough27, dd5, distance, err27	90
		CAS	43.831	0.028	0.354	-7.00E-04	0.028	0.4962	-0.1959	0.00350	p < 0.001	err27, dd5, imperv, distance, elev, map, burn	80

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Between 120	Habitat	FST	37.846	0.024	0.335	1.00E-04	0.0239	0.7774	0.2916	0.00317	p < 0.001	imperv, meadow, canopy, distance, burn, forest, wetland, therm720, water	60
		DPS	46.847	0.029	0.315	-0.0022	0.0295	0.3353	-0.1907	0.00461	p < 0.001	distance, imperv, forest, burn, meadow, water, canopy	70
		DAN	33.598	0.036	0.376	-0.0013	0.0363	0.5621	-0.2137	0.00620	p < 0.001	imperv, forest, water, distance, meadow, burn, canopy	70
		CAS	41.184	0.029	0.365	-0.0013	0.0288	0.4622	-0.3435	0.00365	p < 0.001	imperv, forest, burn, distance, water, meadow	70
	Temp Moist	FST	42.120	0.027	0.375	-2.00E-04	0.0267	0.8003	0.4616	0.00292	p < 0.001	distance, ffp, dd5, gsp, map	70
		DPS	46.027	0.035	0.374	-0.0013	0.035	0.4262	-0.4278	0.00472	p < 0.001	gsp, dd5, distance, map	80
		DAN	31.604	0.043	0.448	-0.0012	0.0433	0.6893	-0.0182	0.00636	p < 0.001	ffp, distance, gsp, dd5, map	80
		CAS	42.901	0.031	0.388	-0.0015	0.0307	0.4643	-0.2615	0.00355	p < 0.001	ffp, gsp, dd5, distance, map	80
	Topo	FST	37.489	0.026	0.368	4.00E-04	0.0263	0.7648	0.3303	0.00317	p < 0.001	elev, err15, rough27, rough15, rough3, ridge	60
		DPS	40.771	0.037	0.39	-6.00E-04	0.0366	0.4489	-0.3006	0.00520	p < 0.001	rough27, distance, elev, err27	90
		DAN	27.817	0.045	0.469	-5.00E-04	0.0454	0.6586	-0.1096	0.00670	p < 0.001	distance, rough27, elev, err27	80
		CAS	36.401	0.033	0.412	-6.00E-04	0.0326	0.3675	-0.3341	0.00394	p < 0.001	elev, rough27, distance, rough15, err27	80

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	43.229	0.022	0.31	1.00E-04	0.0221	0.7543	0.3887	0.00291	p < 0.001	meadow, imperv, dd5, ffp, gsp, map, canopy, burn, elev, distance, err15, ridge, rough27, wetland, forest, rough15, rough3, water, cti	40
		DPS	43.207	0.032	0.341	-0.0012	0.0319	0.3698	-0.1125	0.00497	p < 0.001	distance, rough27, map, imperv, burn, gsp, canopy, ssina, err27, water, meadow	70
		DAN	35.165	0.038	0.398	-0.001	0.0385	0.7131	0.0913	0.00604	p < 0.001	imperv, dd5, rough27, err27, distance, map	80
		CAS	43.534	0.028	0.358	-0.0016	0.0283	0.4551	-0.1857	0.00353	p < 0.001	dd5, distance, imperv, err27, map, burn	90
Between 240	Habitat	FST	32.392	0.027	0.38	5.00E-04	0.0271	0.8336	0.4261	0.00343	p < 0.001	imperv, meadow, canopy, forest, distance, wetland, burn, therm720, water, therm360	40
		DPS	41.850	0.033	0.353	-0.0024	0.0331	0.3888	-0.2576	0.00514	p < 0.001	distance, burn, meadow, forest, canopy, water, imperv	50
		DAN	31.683	0.039	0.406	-0.0011	0.0392	0.6032	-0.1374	0.00638	p < 0.001	distance, imperv, forest, meadow, water, burn, canopy	60
		CAS	39.921	0.03	0.375	-0.0014	0.0297	0.3957	-0.3879	0.00375	p < 0.001	distance, burn, imperv, water, forest, meadow, canopy	60
	Temp Moist	FST	41.684	0.027	0.382	8.00E-04	0.0272	0.9113	0.5757	0.00293	p < 0.001	map, gsp, ffp, distance, dd5, ssina	70

		DPS	44.745	0.037	0.392	-4.00E-04	0.0368	0.3881	-0.4007	0.00482	$p < 0.001$	ssina, map, gsp, distance	80
		DAN	32.635	0.043	0.442	-9.00E-04	0.0427	0.6727	-0.0313	0.00626	$p < 0.001$	ssina, map, ffp, gsp, distance, dd5	70
		CAS	42.204	0.031	0.394	-0.0011	0.0312	0.3469	-0.3795	0.00359	$p < 0.001$	map, ssina, distance, dd5, gsp, ffp	70
	Topo	FST	35.400	0.03	0.418	2.00E-04	0.0298	0.8645	0.4069	0.00327	$p < 0.001$	distance, elev, ridge	90
		DPS	42.289	0.035	0.377	-0.0017	0.0353	0.3968	-0.2021	0.00508	$p < 0.001$	rough27, distance, elev, ridge	80
		DAN	31.494	0.044	0.451	-0.0014	0.0436	0.6685	0.0385	0.00634	$p < 0.001$	rough27, distance, elev, ridge	70
		CAS	37.446	0.033	0.416	-0.0014	0.0329	0.4305	-0.1757	0.00393	$p < 0.001$	distance, rough27, err27	90
	Full	FST	42.160	0.024	0.341	6.00E-04	0.0243	0.7865	0.5715	0.00293	$p < 0.001$	imperv, ssina, rough15, meadow, rough27, ridge, map, ffp, canopy, dd5, distance, wetland, gsp	60
		DPS	44.551	0.032	0.341	-0.0013	0.032	0.5344	-0.1089	0.00485	$p < 0.001$	ssina, map, rough27, canopy, distance, imperv, ridge, meadow, burn	40
		DAN	34.227	0.038	0.396	-1.00E-04	0.0383	0.68	-0.0569	0.00606	$p < 0.001$	ssina, rough27, imperv, map, distance, ridge, canopy	50
		CAS	40.182	0.028	0.36	-0.0011	0.0285	0.4138	-0.2975	0.00374	$p < 0.001$	ssina, rough27, imperv, burn, map, distance, ridge, canopy	50

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Between 480	Habitat	FST	37.685	0.025	0.352	-9.00E-04	0.0251	0.5522	0.0569	0.00316	p < 0.001	imperv, meadow, distance, burn, wetland, canopy, forest, therm720	50
		DPS	41.743	0.032	0.348	-0.0014	0.0325	0.4669	0.0237	0.00507	p < 0.001	meadow, distance, burn, imperv, canopy, forest, water	70
		DAN	29.645	0.04	0.421	-0.0012	0.0406	0.6472	-0.0522	0.00647	p < 0.001	meadow, distance, water, imperv, canopy, burn, forest	50
		CAS	39.937	0.029	0.369	-0.0015	0.0291	0.429	-0.1579	0.00378	p < 0.001	meadow, imperv, distance, water, forest, burn, canopy	60
	Temp Moist	FST	38.618	0.028	0.388	7.00E-04	0.0277	0.8492	0.5744	0.00312	p < 0.001	map, gsp, distance, ffp, ssina, dd5	80
		DPS	43.988	0.035	0.371	-9.00E-04	0.0347	0.4213	-0.3224	0.00490	p < 0.001	ssina, map, distance, gsp, ffp, dd5	60
		DAN	33.202	0.042	0.435	-5.00E-04	0.0419	0.6888	-0.01	0.00616	p < 0.001	ssina, map, ffp, dd5, distance, gsp	70
		CAS	41.494	0.031	0.394	-0.0012	0.0311	0.413	-0.3604	0.00362	p < 0.001	map, ssina, distance, ffp, dd5, gsp	50
	Topo	FST	34.506	0.029	0.414	-2.00E-04	0.0295	0.8031	0.177	0.00330	p < 0.001	elev, ridge, rough27, distance, rough15, rough3	80
		DPS	38.742	0.035	0.376	-0.0016	0.0351	0.5439	-0.1615	0.00532	p < 0.001	distance, rough27, ridge, err3, elev, rough15	80
		DAN	29.445	0.043	0.443	-0.0011	0.0427	0.6927	0.1503	0.00652	p < 0.001	rough27, distance, ridge, elev, rough3, rough15, err27, err3	70
		CAS	35.043	0.032	0.406	-0.0017	0.032	0.4959	-0.1007	0.00404	p < 0.001	distance, rough27, err27, rough15, ridge, elev, err3	80

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	41.411	0.026	0.358	-4.00E-04	0.0256	0.7709	0.4017	0.00297	p < 0.001	meadow, ridge, ssina, imperv, canopy, gsp, rough15, wetland, rough27, dd5, elev, map	50
		DPS	40.332	0.035	0.376	-0.0012	0.0352	0.4478	-0.1472	0.00524	p < 0.001	ssina, canopy, meadow, map, distance, ridge, imperv	40
		DAN	31.884	0.041	0.427	-0.0011	0.0412	0.6734	0.0076	0.00626	p < 0.001	ssina, rough27, map, ffp, ridge, canopy, meadow, distance, imperv	60
		CAS	39.068	0.029	0.369	-0.0016	0.0291	0.4206	-0.3354	0.00373	p < 0.001	ssina, ridge, imperv, canopy, burn, meadow, map, err27, rough27, hsp, insoy, dd5, distance, therm, ffp, scosa, gsp, forest, rough15, elev	30
Between 960	Habitat	FST	42.189	0.024	0.338	-2.00E-04	0.0241	0.5771	0.281	0.00294	p < 0.001	canopy, meadow, wetland, distance, therm720, imperv, water	50
		DPS	41.582	0.034	0.362	-0.0019	0.0339	0.4055	-0.1581	0.00517	p < 0.001	distance, meadow, canopy, burn, forest, water	70
		DAN	30.625	0.041	0.423	-0.0014	0.0409	0.6086	-0.1572	0.00642	p < 0.001	distance, water, meadow, wetland, forest, burn, canopy, imperv	50
		CAS	38.256	0.03	0.376	-0.0013	0.0297	0.4227	-0.2862	0.00389	p < 0.001	meadow, distance, wetland, water, imperv, burn, forest, canopy	60

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Temp Moist	FST	38.116	0.028	0.394	2.00E-04	0.0281	0.8314	0.4672	0.00315	p < 0.001	distance, dd5, map	90
		DPS	44.387	0.035	0.377	-0.0011	0.0354	0.4189	-0.1039	0.00487	p < 0.001	gsp, ssina, dd5, map, distance	70
		DAN	32.907	0.041	0.427	-0.0011	0.0412	0.7008	0.1556	0.00625	p < 0.001	ssina, dd5, ffp, scosa, map, distance, gsp, hli	60
		CAS	42.403	0.031	0.396	-0.001	0.0313	0.4126	-0.1551	0.00359	p < 0.001	dd5, ssina, distance, gsp, map	70
	Topo	FST	34.588	0.031	0.439	-3.00E-04	0.0313	0.7716	0.2006	0.00330	p < 0.001	distance, ridge elev	90
		DPS	41.556	0.038	0.406	-0.0015	0.0381	0.4572	-0.1321	0.00509	p < 0.001	distance, ridge, rough27	90
		DAN	29.812	0.045	0.467	-0.0015	0.0451	0.6921	0.0757	0.00651	p < 0.001	distance, err27, ridge, rough15, elev, rough3, rough27, err16	80
		CAS	38.920	0.034	0.426	-0.0011	0.0337	0.362	-0.389	0.00379	p < 0.001	ridge, distance, elev	90
	Full	FST	46.374	0.024	0.333	4.00E-04	0.0237	0.6964	0.6231	0.00274	p < 0.001	meadow, canopy, ridge, wetland, gsp, dd5	50
		DPS	45.983	0.036	0.39	-9.00E-04	0.0365	0.482	-0.0634	0.00470	p < 0.001	ridge, distance, ssina	80
		DAN	32.840	0.044	0.456	-7.00E-04	0.0441	0.7293	0.3226	0.00622	p < 0.001	dd5, ssina, meadow, ridge	80
		CAS	38.252	0.029	0.371	-9.00E-04	0.0293	0.4218	-0.1582	0.00384	p < 0.001	ssina, meadow, ridge, canopy, dd5, rough3, distance, burn, scosa, rough15, forest, err27, elev, ffp, water, cti, err15, hli, map, rough27, err3, imperv, hsp, gsp, wetland	50

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
Between X	Habitat	FST	39.418	0.024	0.337	2.00E-04	0.024	0.596	0.233	0.00308	p < 0.001	canopy_60, imperv_240, wetland_960, distance, meadow_960, forest, canopy, meadow_480, meadow_120, canopy_120, imperv_60, canopy_480, imperv_120, meadow_60, meadow_240, canopy_60	50
		DPS	42.294	0.03	0.32	-0.001	0.03	0.4051	-0.2045	0.00507	p < 0.001	distance, canopy_240, canopy_480, forest, imperv_240, imperv_120, burn, meadow_960, meadow_480, imperv_480, canopy_120, meadow_240, burn_120, water_120, meadow_120	30
	Temp Moist	FST	41.843	0.026	0.371	2.00E-04	0.0264	0.8456	0.4765	0.00292	p < 0.001	dd5, ssina_480, distance, ffp, gsp_60, map, ffp_60	90
		DPS	44.545	0.035	0.369	-7.00E-04	0.0346	0.3624	-0.2817	0.00488	p < 0.001	map_120, map, ssin_960, ssina_480, ssina_240, gsp, dd5_960, distance	40
	Topo	FST	41.097	0.027	0.384	0	0.0274	0.7475	0.2029	0.00299	p < 0.001	distance, elev, ridge_960, err15_60	80
		DPS	43.774	0.037	0.391	-0.0012	0.0366	0.2922	-0.2597	0.00493	p < 0.001	ridge_960, distance, rough27_240	80

Hypothesis	Class	Dist	R^2	rmse	Rmsd	resd.mean	resd.stdv	resd.skew	Resd.kurt	MSE	Sig	Model	Threshold
	Full	FST	45.324	0.021	0.296	6.00E-04	0.0211	0.6515	0.3118	0.00282	p < 0.001	Canopy_960, ridge_480, dd5_960, ssina_480, gsp_240, meadow_60, ridge_240, meadow_960, ssina_240, wetland_960, meadow_480, ridge_960, gsp_480, imperv_240, meadow_120, gsp_960. imperv_120, err15_60, water_120, imperv_60, meadow_240, canopy_120, canopy_60, err15_120, cti_120, cti_60, canopy, imperv_480, gsp_120, gsp_60	40
		DPS	44.848	0.033	0.349	-5.00E-04	0.0327	0.4108	-0.2252	0.00481	p < 0.001	ssina_240, ssina_480, map, ssina_960, ridge_960, canopy_480, map_240, meadow_960, canopy_240, imperv_240, map_120, imperv_480, imperv_120, dd5_960	60

Appendix 9 Full color versions of Figures 9 and 11 from text.

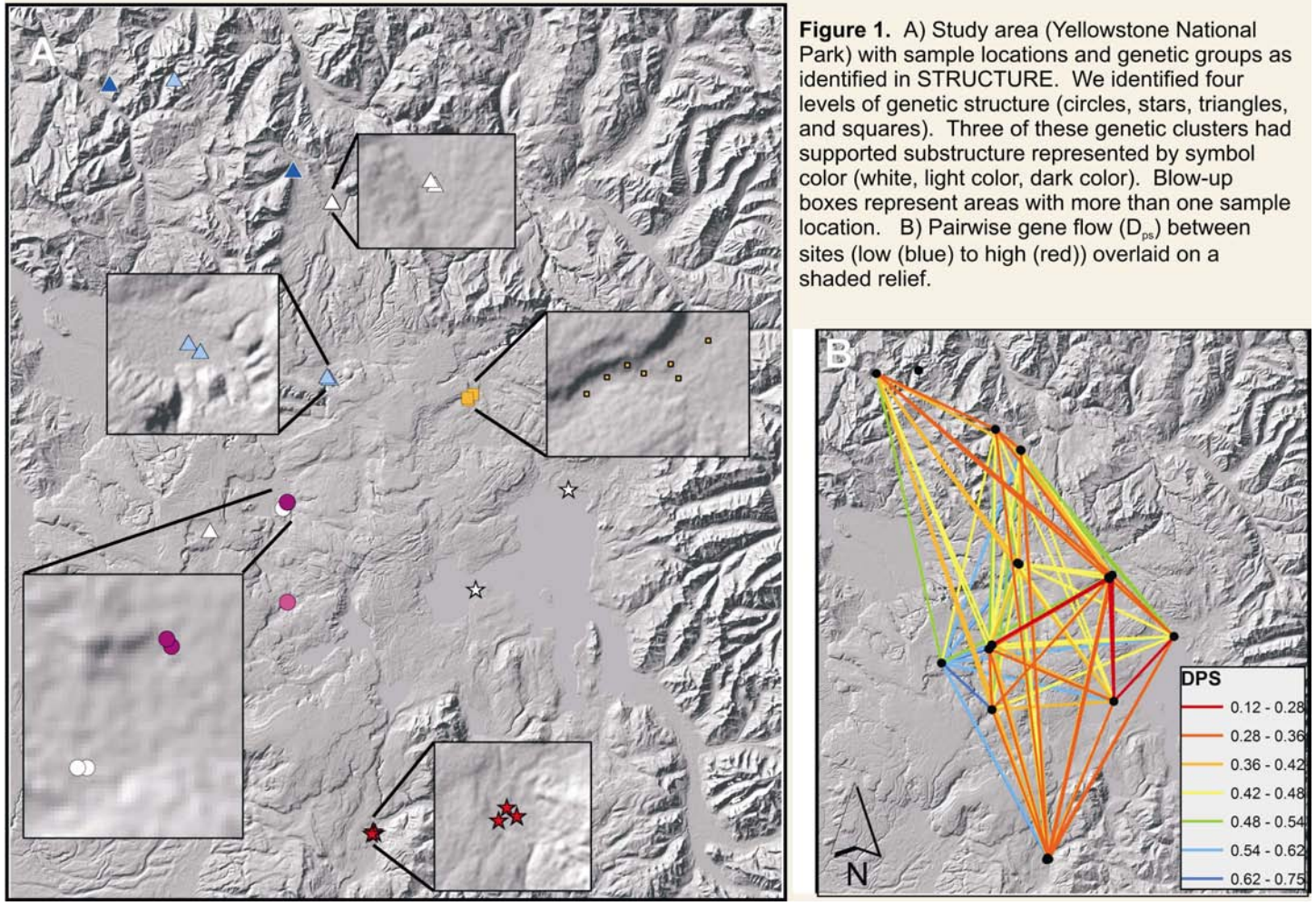


Figure 1. A) Study area (Yellowstone National Park) with sample locations and genetic groups as identified in STRUCTURE. We identified four levels of genetic structure (circles, stars, triangles, and squares). Three of these genetic clusters had supported substructure represented by symbol color (white, light color, dark color). Blow-up boxes represent areas with more than one sample location. B) Pairwise gene flow (D_{ps}) between sites (low (blue) to high (red)) overlaid on a shaded relief.

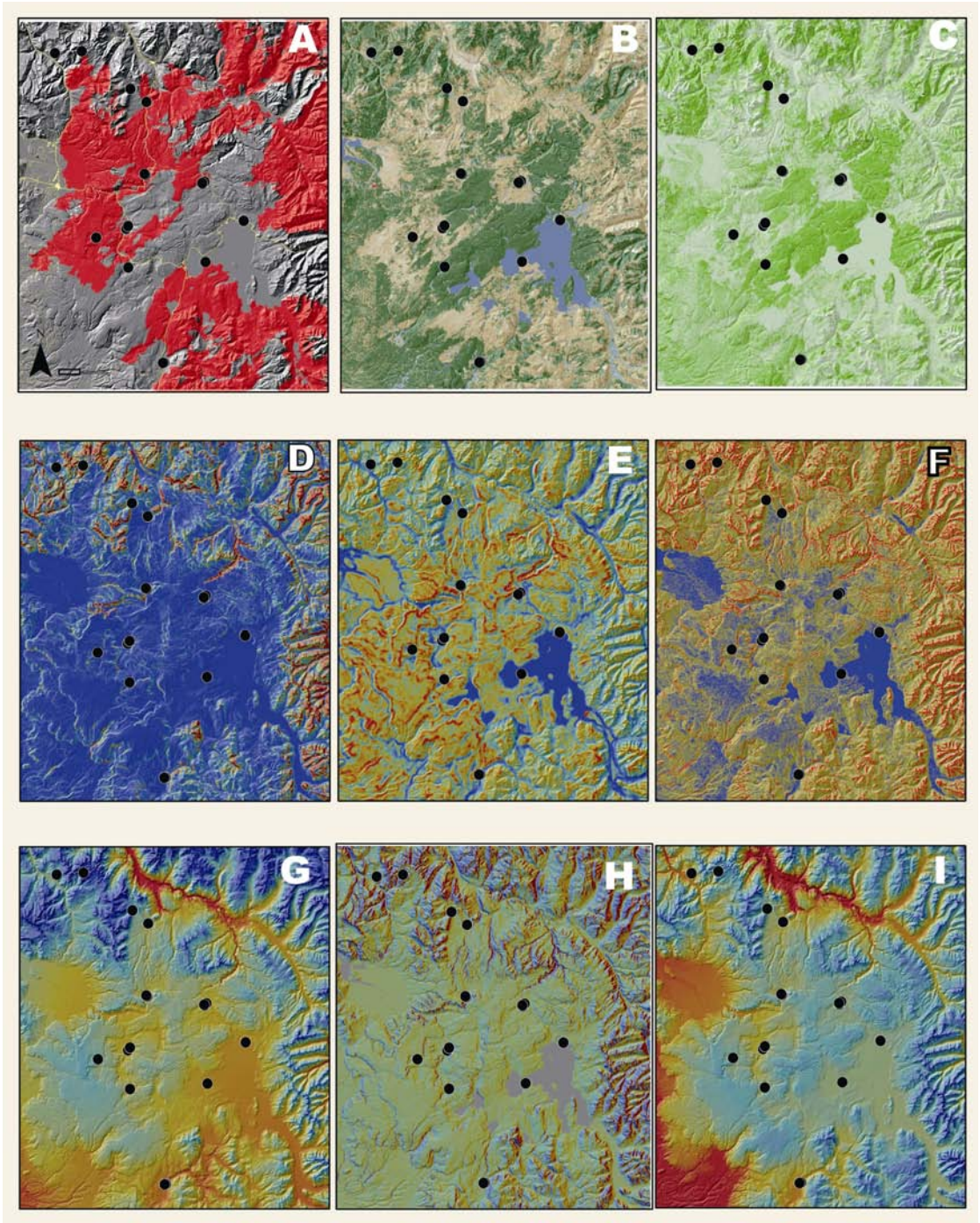


Figure 11. Select important metrics mapped for the study area by ecological process (habitat permeability (A-C), ecological morphology (D-F), and temperature-moisture (G-I)). Circles represent sample locations. Metrics are as follows: A) 1988 fire perimeters (red) with impervious surfaces (yellow), B) NLCD classification, C) canopy cover (0 (lt. green)- 100%

(dark green)), D) roughness at 15X15 window (0 (blue)- 100% (red)), E) elevation relief ratio (0 (blue)- 100% (red)), F) hierarchical slope position (blue – orange, ridges in red), G) growing season precipitation (low (red)- high (blue)), H) ssina (0 (blue) – 100% (red)), and I) dd5 (low (blue) – high (red)).

CHAPTER 4 - Frogs in the high mountains: application of gravity models for landscape genetics

Melanie A Murphy, Ray Dezzani, David Pilliod, & Andrew Storfer

Melanie A Murphy (corresponding author) (mamurphy@wsu.edu) School of Biological Sciences, Washington State University, Pullman WA, 99164. Ray Dezzani (rdezzani@uidaho.edu), Department of Geography, University of Idaho, Moscow, ID 83844. David Pilliod (dpilliod@usgs.gov), USGS Forest and Rangeland Ecosystem Science Center, Snake River Field Station, Boise Id 83706. A Storfer (astorfer@wsu.edu), School of Biological Sciences, Washington State University, Pullman WA, 99164.

Keywords: landscape genetics, gravity model, connectivity, gene flow, metapopulation, amphibians

For submission to The American Naturalist

For expanded on-line appendices:

Methods: Multilocus genotyping and validation (Appendix 10)

Gravity model justification and derivation (Appendix 11)

Expanded version of table 8 (Appendix 12)

ABSTRACT

Landscape genetics, which focuses on quantifying the impact of landscape composition on gene flow, is ideally suited for addressing metapopulation connectivity. Metapopulation theory highlights the importance of landscape composition and dispersal for maintaining population persistence. In a discrete patch system such as typically occupied by pond breeding amphibians, landscape properties both at and between sites may affect population connectivity. Therefore, we adapt gravity equations to predict genetic connectivity as a function of both at site and between site landscape processes. We then apply gravity models to Columbia spotted frogs (*Rana luteiventris*) in the Bighorn Crags, Idaho USA (8 loci, 37 sites, n=441). Using standard analyses, we found significant isolation-by-distance, multiple levels of genetic clustering (STRUCTURE), and more variation explained between basins than between drainages (AMOVA). The gravity model allowed us to reveal additional information: population connectivity is correlated with distance, at site landscape processes (predation, site productivity), and between site landscape processes (major topographic features, frost-free period). The negative effect of predation and positive effect of site productivity on connectivity in combination with bottleneck tests allowed us to infer source-sink dynamics. Finally, gravity models prove to be a highly powerful tool to predict metapopulation connectivity in systems with discrete habitats.

Keywords: Landscape genetics, gravity models, metapopulations, source-sink, *R. luteiventris*, amphibians

INTRODUCTION

A central objective in biology and conservation is to understand species' connectivity. Connectivity is central in metapopulations dynamics where migration of individuals among subpopulations is critical for recolonization and long-term viability (Bouchy et al. 2005; Jiang et al. 2007; Ovaskainen and Hanski 2004; Wiens 1996). Pond breeding amphibians are often thought to function as metapopulations (Compton et al. 2007; Smith and Green 2005), with an inhospitable habitat matrix among sites resulting in restricted connectivity and elevated extinction risk (Storfer 2003; Wiens 1996). Understanding how landscape processes affect connectivity in the context of metapopulation structure is critical for conservation as many amphibian species are in global decline, with habitat fragmentation and alteration a primary cause (Collins and Storfer 2003; Stuart et al. 2004). The affect of reduction in the proportion of suitable habitat across the landscape may be difficult to quantify (Gardner et al. 2007), particularly in metapopulations where habitat modifications may have a non-linear cumulative effect on dispersal and recolonization (Gilpin 1991).

Collecting sufficient demographic data to estimate landscape effects on connectivity among subpopulations can be problematic, particularly in amphibian populations that fluctuate widely in size from year to year (Green 2003; Pechmann et al. 1991). Alternatively, connectivity among sites can be estimated by genetic distance based on neutral genetic markers such as microsatellites (Keyghobadi et al. 2005). Landscape genetics is a newly emergent field combining population genetics and landscape ecology well suited to address these effects of landscape process on population genetic connectivity (Manel et al. 2003; Storfer et al. 2007).

In metapopulations, amount of connectivity is likely driven by both between and at site processes (Hanski and Gaggiotti 2004). Between sites, distance and the suitability of intervening

landscape matrix for dispersal will influence gene flow and recolonization potential (Beebee 2005; Manier and Arnold 2006; McRae 2006; Pellet et al. 2007). At sites, productivity (Johnson and Semlitsch 2003) and predation (Bosch et al. 2006; Orizaola and Brana 2006; Pilliod and Peterson 2001) control the number of offspring produced driving source-sink dynamics (Pulliam 1988). The more offspring produced, the more potential for connectivity between sites and likelihood of being a source population (Pulliam 1988). In addition, at site characteristics may make a location both attractive to potential migrants and sustain those arriving individuals.

Gravity models (Fotheringham and O'Kelly 1989), or spatial interaction models, are well-suited for assessing metapopulation connectivity whereby flow is governed by both at site and among site processes. Based on Newton's law of gravitation, gravity models explicitly model flow or number of trips based on "mass" or potential flow from a site (Anderson 1979). Gravity models are typically used to for predicting transportation flow and trade of economic goods (Voorhees 1956; Willig and Bailey 1979). Application in natural sciences is relatively new, with examples in epidemiology (Xia et al. 2004) and ecology (Bossenbroek et al. 2007; Bossenbroek et al. 2001; Ferrari et al. 2006) estimating connectivity and spread. In addition to the ability to include at site data, gravity models have advantageous model properties compared to matrix approaches typically applied in landscape genetics (Mantel 1967). Estimates are multivariate and do not suffer from potential bias in assessing model uncertainty (Castellano and Balletto 2002; Legendre 2000; Raufaste and Rousset 2001; Rousset 2002).

The Columbia spotted frog (*Rana luteiventris*) functions as metapopulations in the Bighorn Crags, Idaho with processes influencing breeding site occupancy both at sites and between sites (Pilliod et al. 2002). Located in the Frank-Church Wilderness Area, the system has extreme climatic and topographic conditions represented by a limited growing season and granitic peaks

over 3000 m separating basins (Pilliod et al. 2002). In this high elevation system, differences in breeding phenology between low elevation and high elevation sites may result in a biological barrier to gene flow (Palo et al. 2003; Giordano et al 2007). In addition, introduced trout species prey on *R. luteiventris* eggs, larvae and newly metamorphosed individuals in otherwise suitable habitat (Pilliod and Peterson 2001) which may result in a population sink. Despite the observed ability to move though inhospitable habitat matrix (Pilliod et al. 2002), sites in the southern basins contain significant population genetic structure suggesting little connectivity among sites (Funk et al. 2005). By including both at site and between site landscape processes, we can estimate fine-scale metapopulation connectivity of *R. luteiventris* in a natural environment.

Using traditional population genetic approaches and newly derived gravity models, we investigated fine-scale landscape genetic structure of *R. luteiventris* in eight basins in the Bighorn Crags, Idaho. Our goal was to model gene flow among sites as a function of three types of independent variables: distance between sites, landscape processes at sites, and landscape processes between sites. We addressed three primary questions: i) is genetic structure clustered by basin or drainage?, ii) is connectivity influenced by both at site and among site landscape processes?, and iii) which landscape processes explain observed connectivity among sites? Based on previous research identifying ridges as potential barriers to gene flow in amphibians (Funk et al. 2005; Lougheed et al. 1999), we hypothesize that more genetic variation can be explained by basin than by drainage. Because at site characteristics influence the number of potential migrants, we hypothesize both at site and among site landscape processes significantly influence *R. luteiventris* connectivity. We predict high connectivity between sites will be associated with high primary productivity and lack of fish. In addition, areas of high connectivity will have similar elevation, higher than average rainfall, and a longer growing

season than areas of low connectivity. Finally, connectivity will be limited by major topographic features and geographic isolation.

METHODS

Study area and field methods

We collected tissue samples from *R. luteiventris* at 37 sites in eight basins in the Bighorn Crags in 2005-06 (Fig. 12). Field crews collected buccal swabs from adults (Goldberg et al. 2003) and a small tail clip (~2 mm) from larvae with the goal of at least 20 individuals per site. We also collected data on surveyed wetlands including: wetland size (area, perimeter, depth), distance to habitat types (forest, shrub, rock), and presence of fish (Pilliod et al. 2002).

(Figure 12 caption)

Figure 12 shows the study area with sites (A), pruned network (B), and location overview (lower right hand corner). A) Figure 12A is a map of study area with basins labeled in text. Major genetic groups identified in STRUCTURE (Table 7), are shown as circles (I) and triangles (II). Subclusters are shown by gray scale (I: black(A), gray (B), and white (C) and II: black (A) and gray (B)). Symbols with a black dot in the center represent sites where assignment was nebulous, likely due to small sample size (Table 8). B) Figure 12B shows the pruned network including only sites with large enough sample size for reliable pair-wise estimates of gene flow.

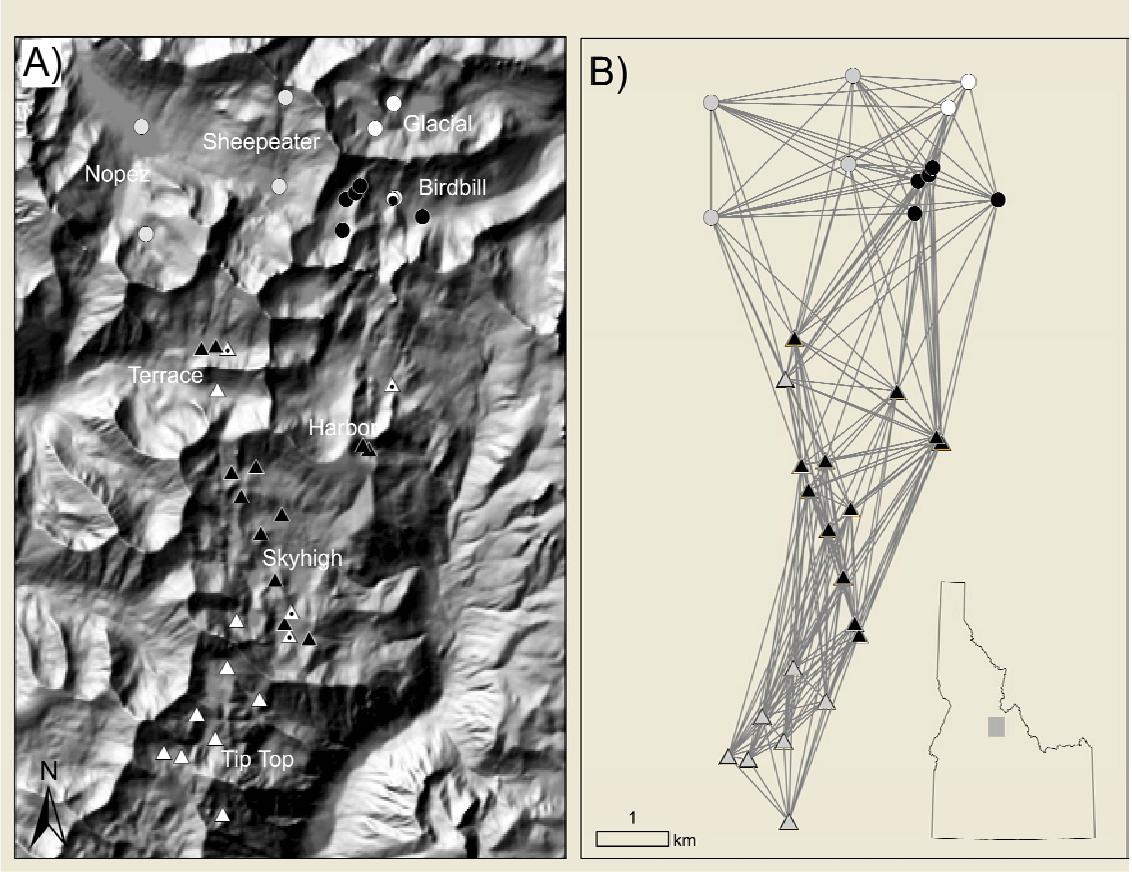


Figure 12

Genetic clustering

DNA was extracted using the Qiagen DNeasy96 tissue protocol (Qiagen Inc). We generated multi-locus genotypes using eight microsatellite loci (Funk et al. 2005; Monsen and Blouin 2003), an ABI 3730 automated sequencer and scored with GeneMapper 3.7 (ABI) (for loci, optimized conditions, quality control, and basic validation see Appendix 10). We implemented a maximum likelihood algorithm (COLONY) to identify full sibling larvae (Wang 2004). We then subsampled by site at the size of the smallest sibling cluster to avoid biasing allele frequency distributions (C. Goldberg, pers. comm.). We tested all loci for concordance with Hardy-

Weinberg equilibrium and significant linkage disequilibrium between all pairs of loci using FSTAT 2.9.3.2 and GENEPOP 3.4 (Goudet 2001; Raymond and Rousset 1995).

We estimated gene flow and other standard measures of population structure and genetic distance with GENEPOP 3.4 (Raymond and Rousset 1995) and Microsatellite Analyser (Dieringer and Schlötterer 2003). We tested overall isolation-by-distance by calculating a Mantel statistic (Mantel 1967) between genetic distance (Nei 1972) and geographic distance in GenAlEx 6.0 (Peakall and Smouse 2006). We assessed overall hierarchical partitioning of genetic structure by site, basin, and drainage using an Analysis of Molecular Variation (AMOVA) with the following parameters: co-dominate markers, F_{ST} as the distance statistic, and 999 permutations (Peakall and Smouse 2006; Peakall et al. 1995). We compared three hierarchical population groupings with AMOVA to test which explained the most genetic variation: site – basin, site – drainage, and basin – drainage.

We assessed the clustering of multilocus genotypes with no *a priori* definition of population identification using program STRUCTURE (Prichard et al. 2000) with 500,000 iterations burn-in, 200,000 Markov-chain Monty-Carlo replicates post burn-in, correlated allele frequencies, admixture model, inferred alpha with uniform prior, and number of populations (*K*) for 1-20. We estimated major genetic clusters (*K*) using the methods proposed by Evanno et al. (2005). Because this approach identifies the first major break in genetic variation, we reran each identified genetic cluster in STRUCTURE until *K*=1 had the most support to identify hierarchical substructure. To visualize connectivity, we estimated the most likely connections among sites using a graph theory based maximum likelihood approach (Dyer and Nason 2004). Finally, to identify potential population sinks, we assessed bottleneck effects as identified by sign-ranked (models: IAM, SSM), Wilcoxon (models: IAM, SSM), and mode-shift tests in program

BOTTLENECK (Cornuet and Luikart 1997; Luikart and Cornuet 1998; Luikart et al. 1998a; Luikart et al. 1998b).

Landscape genetics models

Introduction to gravity models

Gravity models fit our data structure with "trips" (T_{ij}, number of trips between sites *i* and *j*, in our case gene flow) estimated by three model components: weights (w_j, distance to destination site), production/attraction measures (v_i, landscape processes at site - the "mass" of the gravity equation), and resistance measures (c_{ij}, landscape processes between sites) (Fotheringham and O'Kelly 1989). The gravity equation takes the following form with μ, α, and β as the estimated parameters (eq. 1, Fotheringham and O'Kelly 1989):

$$T_{ij} = k v_i^{\mu} w_j^{\alpha} c_{ij}^{-\beta} \qquad \text{(eq. 1)}$$

Samples at a node share a proportion of co-dominant alleles from neutral loci (as measured by D_{ps}, Bowcock et al. 1994) such that the role of intervening landscape can be used to predict the spatial distribution of gene flow (Appendix 11). Thus, we can estimate "flow", performing the equivalent of estimating a trip distribution matrix from transportation to allocate flows given a network of sites (Anderson 1979; Fotheringham and O'Kelly 1989).

We estimated landscape genetic structure using the above gravity equation (eq. 1). We measured flow between sites (T_{ij}) as $1-D_{ps}$ (Bowcock et al. 1994) because allelic similarity in allele frequency distributions can be justified as a substitution for the number of trips (Appendix 11), this genetic distance measure has relaxed equilibrium assumptions compared to $F_{ST,}$ and it has the power to detect fine-scale genetic structure (Bowcock et al. 1994; Murphy et al. submitted). A more detailed justification and derivation of gravity models for landscape genetic application is found in Appendix 11.

Network optimization and bandwidth

In transportation applications of gravity models, the network on which flow is measured is generally a physical entity such as road network. In this case, all the between site connections and spatial locations of those connections are known. In our landscape genetic application, we must estimate both the network and an area of influence of between site connections ("bandwidth"). To estimate the network, we first connected each site to every other site ("saturated network"). However, we know that this is not a realistic representation of direct gene flow. However, we did not want to optimize the network based on the dependent variable (e.g., a threshold gene flow value) or fit the network to independent variables (e.g., drop connections based on model fit). Both these approaches may introduce bias when inferring process based in part on the network topology. However, direct gene flow will not occur past the maximum dispersal distance of the species. Therefore, we estimated a gravity model for the saturated network and for a network pruned (i.e., connections dropped) to include pair-wise connections less than 4000 m (~2 times the maximum recorded travel distance for *R. luteiventris* in the Bighorn Crags (Pilliod et al. 2002)). We optimized connection bandwidth by testing 30, 120, 240, 480, and 960m sections of landscape between sites. We selected the final bandwidth based on overall model fit.

Independent variables

For the distance term (w), we calculated topographically corrected distance between sites. We also estimated landscape processes influencing production/attraction (v) at the site and resistance (c) between sites (Table 7, see Appendix 12 for detailed variable explanation, source data, predicted effect, and ecological justification).

Production/attraction of migrants may be influenced by presence of fish, site productivity, water permanence, and quality of surrounding habitat. To include a measure of predation risk, we incorporated wetland depth as deeper wetlands tend to have introduced trout predators (*Oncorhynchus clarki*, *O. mykiss*) (Pilliod et al. 2002). Primary productivity and water temperature, as measured by solar radiation (McCune and Keon 2002) and elevation, are likely limitations to larval development at high elevations (Palo et al. 2003; Giordano et al 2007). Site permanence may also be important as temporary water bodies may evaporate before larva metamorphose. Compound topographic index measures the potential hydrological flow into an area and water holding capacity, thus is a measure of site permanence (Moore et al. 1993). In addition, surrounding habitat may provide cover and foraging for breeding adults and newly metamorphosed individuals. To assess dominant cover around a site, we measured the distance from the site to forest, rock, and shrub habitat (Table 7, Appendix 12).

We assessed landscape resistance between sites by habitat permeability, topographic morphology, and temperature-moisture regime with a set of variables derived from 30m data (for complete list of variables see Table 7, Appendix 12). We assessed habitat permeability by cover classes shown to have an impact amphibian connectivity other studies: forest (Eigenbrod et al. 2008), meadow (Munger et al. 1998), and water (Munger et al. 1998; Pilliod et al. 2002) (Table 7; Appendix 12). Topography may restrict gene flow by either acting as a physical barrier (Funk et al. 1999; Lougheed et al. 1999) or due to phenological differences between low and high elevation sites (Funk et al. 2005; Giordano et al. 2007). We measured topographic morphology with a set of DEM derived variables: change in elevation between sites, ridgelines, elevation relief ratio (Evans 1972), relative slope position (Murphy et al. submitted), and hierarchical slope position (Murphy et al. submitted) (Table 7, Appendix 12). Amphibians generally have little

physiological control over water loss (Bartelt and Peterson 2005; Duelman and Trueb 1994). Therefore, temperature-moisture regime may directly affect connectivity (Pilliod et al. 2002). We calculated average temperature-moisture condition between sites based on both topography (compound topographic index and heat load index) and climate (temperature and rainfall, Table 7).

Parameter	Process	Variable	Code
Distance (w)	Isolation by distance	Topographic Distance	Distance
Production (v)	Productivity	Elevation	Elev
		Heat load index	Hli
		Compound topographic index	cti
	Predation	Depth	Depth
Resistance (c)	Habitat	Water	Water
		Forest	forest
		Meadow	Meadow
	Topo-morph	Elevation relief ratio[1]	Err
		Relative slope position	Rsp
		Hierarchical slope position	Hsp
		Ridges	Ridge
	Temp-Moist	Heat load index	Hli
		Compound topographic index	cti
		Precipitation ratio	Pratio
		Frost free period	Ffp

Table 7

Independent variables used to build gravity models of metapopulation connectivity.

The table headings are as follows: Parameter - the gravity model parameter estimated by a given set of independent variables (see eq. 1), Process - the landscape process being measured for a set of independent variables (Topo-morph – topographic morphology, Temp-Most – temperature-moisture), Variable - the variable name and code is the abbreviation for the given variable (using previously published codes when available).

[1]Elevation relief ratio was calculated at multiple window sizes (3X3, 15X15, and 27X27 30m cells). For an expanded version of this table with variable calculation, references, predicted effects, and ecological justification see Appendix 12.

Estimating gravity models

We estimated global connectivity of *R. luteiventris* in the Bighorn Crags using an unconstrained gravity model (Fotheringham and O'Kelly 1989). As an unconstrained model (i.e. estimate is not distributed based on origin or destination site), the gravity equation (eq. 1) can be linearized as (eq. 2, Fotheringham and O'Kelly 1989):

$$\ln T_{ij} = \ln k + \ln \mu v_i + \ln \alpha w_j - \ln \beta c_{ij} \quad \text{(eq. 2)}$$

We then estimated the parameters using linear regression in R (R Development Core Team 2007). We evaluated model fit using residual maximum likelihood estimation (REML) and parameter significance. For model selection, we implemented the Akaike Information Criterion (AIC) (Aikaike 1973).

To get a more spatially distributed estimate of flow incorporating local effects, gravity models can be either singling constrained (by origin (production, landscape processes influencing flow coming from a site) or destination (attraction, landscape processes influencing flow arriving at a site)) or doubly constrained (by origin and destination sites). We estimated spatially distributed connectivity of *R. luteiventris* in the Bighorn Crags using a gravity model singly constrained by origin (Fotheringham and O'Kelly 1989). The production constrained gravity equation can be linearized as (eq. 3, Fotheringham and O'Kelly 1989):

$$\ln T_{ij} = \ln k_i + (\ln \mu v_i + \ln \alpha w_j - \ln \beta c_{ij}) \quad \text{(eq. 3)}$$

This linearized form is identical to the unconstrained equation (eq. 2), with the exception of a unique constant for each origin site. We estimated parameters using hierarchical linear mixed models grouped by origin site in R (lme function, R Development Core Team 2007). We then evaluated model fit using REML and parameter significance. For model selection, we used AIC (Aikaike 1973).

To validate our gravity models, we evaluated if the spatial distribution of sites explained observed connectivity, overall model significance, and the effect of missing sites on model estimates. To test if observed connectivity is simply a function of the spatial distribution of sites (Rickets 2001), we calculated maximized flow (entropy) through the network (Brooks and Wiley 1988). We then correlated the expected distribution of flows based on entropy with observed gene flow of *R. luteiventris* in the landscape. If distance, at site landscape processes and between site landscape processes have no effect then the observed genetic distances will be highly correlated with entropy expectations. To quantify the probability of getting the selected model by random chance, we randomized observed flow through the network (n=1000), calculated REML, and used this to create a null distribution. We then calculated whether the observed REML was greater than the 95^{th} percentile of this null distribution ($p < 0.05$). We assessed the impact of missing sites by a jackknife of the data, removing one site at a time (and all related pair-wise connections) and re-estimating the model.

RESULTS

Genetic Structure

We found more genetic structure in *R. luteiventris* was explained by basins than drainages. We collected 477 *R. luteiventris* samples (adults (418), tadpoles (59)) and successfully genotyped 442 samples for at least 6 of the 8 loci (Table 8) (Funk et al. 2005; Monsen and Blouin 2003). The northern basins had fewer occupied sites than the southern basins, many with low occupancy as reflected in our sample sizes (Table 8). Overall average heterozygosity across all sites is 0.42, consistent with findings in Funk *et al.* (2005) (Appendix 10). No loci were out

of linkage equilibrium after correcting for multiple tests and no sites with greater than five samples were out of Hardy-Weinberg equilibrium.

F_{ST} between sites ranged from 0.012 – 0.500, with the most genetically divergent sites located in Tip Top basin and the most genetic connectivity in Skyhigh basin (Fig. 12A). Across the study area, we found most sites to be significantly differentiated, although some F_{ST} values were not significant from zero (Appendix 10). However, all pair-wise F_{ST} values between drainages and between basins were significant from zero. Based on the Mantel test between geographic distance and Nei's distance, we found significant isolation-by-distance (Mantel r = 0.20, p-value < 0.001, Appendix 10). Across all the AMOVA tests (Table 9), we found most variation explained within sites. However, grouping by basin minimized more within group variation (5%) and maximized among group variation (12%), compared to drainage (9% within, 10% among) or basin within drainage (7% within, 8% among) (Table 9). All grouping were significant in the AMOVA analyses ($p < 0.001$). Using PopGraph, we were able to visualize genetic clustering by basin based on the most likely connections between sites (Appendix 10).

We identified two major genetic clusters using STRUCTURE, each with significant sub-clustering (Fig. 12A). Main clustering corresponds to a major ridgeline, resulting in a north (I)-south (II) break in the study area. Substructure is by basins, suggesting that topography considerably restricts connectivity (Fig. 12A). We identified three genetic subgroups in the north and two primary subgroups in the south. Of the northern sites with large enough sample sizes to run bottleneck tests, 60% (6/10) had significant bottleneck signatures (heterozygosity excess and/or shifts in allele frequency distribution) for at least 3/5 tests. In contrast, this was true for only one site in the southern cluster (1/26; Appendix 12).

Table 8

Sample summary. Table 8 is a summary of samples collected in the Bighorn Crags including drainage and basin (see Fig. 12 for basin locations). Samples are the number of samples collected as: # adults (# tadpoles). Genotypes are the number of samples successfully genotyped. For an expanded version of the table including bottleneck test results, see Appendix 12.

Drainage	Basin	Site	Cluster	Genotypes	Basin Total	Drainage Total
Clear	Birdbill	Birdbill Lake[1]	IA	4		
		Gentian Lake	IA	20		
		Gentian Ponds	IA	18		
		Meadow Lake	IA	20		
		Mirror Lake[1]	I*	1		
		Stocking Cap Lake	IA	6	69	
	Glacier	Golden Lake	IB	6		
		Pothole Lake	IB	13	19	88
Nopez	Nopez	Nopez Lake	IC	5		
		Ship Island Lake	IC	18	23	
	Sheepeater	Airplane Lake	IC	22		
		Elenas Lake	IC	20	42	65
Waterfall	Terrace	Barking Fox Lake	IIB	14		
		Terrace Lakes[2]	IIA	10	24	24
Wilson	Harbor	Bob Lake	IIA	17		
		Buteo Meadow[1]	II*	4		
		Tobias Lake[1]	IIA	3		
		Welcome Lake	IIA	15	39	
	Skyhigh	Bachelor Meadow	IIA	8		
		Buck Lake[1]	IIA	3		
		Cache Lake	IIA	7		
		Doe Lake	II*	6		
		Egg White Lake	IIA	17		
		Fawn Lake	IIA	17		
		Frog Pond Lake	IIA	10		
		Glacial Lake[1]	IIA	2		
		Homer Pond[1]	II*	1		
		In and Out Lake	IIA	10		
		Skyhigh Lake	IIA	23		
		Mount Wilson	IIA	15		
		Twin Cove Lake	IIA	7	126	
	Tip Top	Greggs Lake	IIB	20		
		Moose Lake	IIB	18		
		Paragon Lake	IIB	20		
		Ramshorn Lake	IIB	18		
		U. Paragon Wetland	IIB	6		
		Walkabout Lake	IIB	17	99	264
Total				441		

[1]Not included in pair-wise analysis (F_{ST} or gravity models) due to small sample size.. [2]Terrace lakes consist of four small interconnected lakes, all with low sample size. Sites were combined for all pair-wise analyses. *Assignment for this site is unclear, likely due to small sample size.

Table 9

AMOVA summary tables. Table 9 is a summary of AMOVA results, testing alternative grouping criteria. Heading are as follows : Groups – tested grouping levels. Df – degrees of freedom. SS – sum of squares, MS – mean sum of squares, Est. Var. – estimated variance, variation – variation explained, p-value – significance of that group.

Summary AMOVA Table – Drainage vs. Site

Groups	Df	SS	MS	Est. Var.	Variation	p-value
Among Drainages	3	137.208	45.736	0.225	10%	<0.001
Among Sites	31	197.427	6.369	0.190	9%	<0.001
Within Sites	843	1483.648	1.760	1.760	81%	<0.001
Total	877	1818.282	53.865	2.175		

Summary AMOVA Table – Basin vs. Site

Groups	Df	SS	MS	Est. Var.	Variation	p-value
Among Basins	7	213.162	30.452	0.244	12%	<0.001
Among Sties	27	121.472	4.499	0.117	5%	<0.001
Within Sites	843	1483.648	1.760	1.760	83%	<0.001
Total	877	1818.282	36.711	2.121		

Summary AMOVA Table – Drainage vs. Basin

Groups	Df	SS	MS	Est. Var.	Variation	p-value
Among Drainages	3	137.208	45.736	0.175	8%	<0.001
Among Basins	4	75.954	18.989	0.155	7%	<0.001
Within Basins	870	1605.120	1.845	1.845	85%	<0.001
Total	877	1818.282	66.570	2.175		

Landscape genetic models

We found both at site and between site landscape processes influenced connectivity, supported by significance of all components of the gravity model (distance, production/attraction, and resistance, p-value < 0.05). The optimal bandwidth, as determined by model fit, was 30m. However, the area of landscape influence considered between sites had little effect on model fit. Using the saturated network, global connectivity was a function of distance,

production (site depth, heat load index), and landscape resistance (frost free period, elevation relief ratio at a 27X27 cell window size; REML = 287.41, AIC = -558.8, Table 10). The next most likely model included relative slope position, but did not improve AIC (Table 10). Pruning the network had little effect on model fit, with all three gravity components still significant (p-value < 0.05). For the pruned network, global connectivity was a function of distance, production (heat load index), and landscape resistance (frost free period, elevation relief ratio at a 27X27 cell window size; REML = 139.3, AIC = -264.6; Table 10).

The constrained gravity model results were very similar. We found all components of the gravity model to be significant. Using the saturated network, global connectivity was a function of distance, production (heat load index), and landscape resistance (frost free period, elevation relief ratio at a 27X27 cell window size; REML = 287.41, AIC = -558.8, Table 10). The same variables were selected using the pruned network (REML = 150.22, AIC = -286.44, Table 10). Although the REML and AIC scores can not be compared between pruned and unpruned networks, unconstrained and constrained models of the same network topology can be compared. Overall, the constrained gravity models incorporating local effects had higher REML and AIC scores, suggesting that singly constrained models are a better fit to the data. All final unconstrained and constrained gravity models were significant compared to null flow (entropy) and randomized D_{ps} (p-values < 0.05). In addition, the jackknife did not change parameter selection and had only trivial affect on parameter estimates (data not shown).

Table 10

Model results. Table 10 displays model results as follows with the selected model in **bold** text: Type – type of model (null entropy, unconstrained gravity model or constrained gravity model), Network – saturated or pruned network, Parameters – selected parameters for each model (dist – distance (w), v – production term, c - resistance term). Test statistics are: residual maximum likelihood (REML) and Akaike Information Criterion (AIC). Variable codes are found in Table 1. Model significance values is the result of the randomization test.

Type	Network	Parameters	REML	AIC	P-value
Entropy	Saturated	None	Na	Na	0.857
	Pruned	None	Na	Na	0.881
Unconstrained	Saturated	Dist	218.90	-447.59	<0.0001
		Dist+ V(Depth)	236.54	-463.08	<0.0001
		Dist + v(hli)	241.52	-473.03	<0.0001
		Dist + c(ffp)	262.97	-515.95	<0.0001
		Dist + v(hli) +c(ffp)	274.92	-537.84	<0.0001
		Dist + v(Depth) + v(hli) + c(ffp)	278.60	-543.20	<0.0001
		Dist+ v(Depth) + v(hli) + c(ffp) + C(rsp)	281.78	-547.56	<0.0001
		Dist + v(Depth) + v(hli) + c(ffp) + c(err27)	**287.41**	**-558.84**	**<0.0001**
		Dist + v(Depth) + v(hli) + c(ffp) + C(err27) + c(rsp)	287.69	-557.38	<0.0001
		Dist + v(Depth) + v(hli) + c(dd5) + C(err27)+ c(scosa)	283.24	-548.47	<0.0001
	Pruned	Dist	97.56	-187.11	<0.0001
		Dist + v(Depth)	108.98	-207.96	<0.0001
		Dist + v(hli)	116.29	-222.59	<0.0001
		Dist + c(dd5)	121.30	-232.60	<0.0001
		Dist + v(Depth) + v(hli)	112.73	-213.46	<0.0001
		Dist + v(hil) + c(ffp)	122.23	-254.46	<0.0001
		Dist + v(hli) + c(ffp)+ c(err27)	**139.30**	**-264.60**	<0.0001
		Dist + v(hli) + c(ffp) + c(rsp)	132.30	-250.59	<0.0001
Constrained	Saturated	Dist	280.41	-552.81	<0.0001
		Dist+ v(Depth)	280.06	-550.12	<0.0001
		Dist + v(hli)	283.09	-556.19	<0.0001
		Dist + c(ffp)	304.09	-598.18	<0.0001
		Dist + v(hli) + c(ffp)	304.16	-594.32	<0.0001
		Dist + v(hli) + c(ffp) +c(rsp)	309.40	-604.80	<0.0001
		Dist + v(hli) + c(ffp) + c(err27)	**311.04**	**-608.09**	<0.0001
		Dist + v(hli) + c(ffp) +c(rsp) + c(elev_diff)	307.09	-598.18	<0.0001
		Dist + v(hli) + c(ffp) +c(rsp) + c(err27)	312.17	-608.34	<0.0001
		Dist + v(Depth) + v(hli) + c(ffp) + c(err27)	309.69	-603.37	<0.0001
	Pruned	Dist	123.78	-257.56	<0.0001
		Dist + v(hli)	127.01	-264.01	<0.0001
		Dist + v(Depth)	133.76	-255.52	<0.0001
		Dist + c(ffp)	141.37	-272.74	<0.0001
		Dist + v(hli) + c(ffp)	144.89	-277.78	<0.0001
		Dist + v(hli) + c(ffp) + c(err27)	**150.22**	**-286.44**	<0.0001
		Dist + v(hli) + c(ffp) + c(err27) +c(rsp)	148.58	-281.15	<0.0001

DISCUSSION

Connectivity and source-sink dynamics are critical components for understanding metapopulation systems (Rickets 2001; Wiens 1996). Using a new implementation of gravity models, we were able to estimate connectivity in a metapopulation and identify potential drivers of source-sink dynamics (Pulliam 1988). Gravity models are unique compared to existing methods in landscape genetics. They are able incorporate both at site and between site landscape processes, directly model flow, and can simultaneously estimate multiple parameters.

Genetic Structure

We found basins, and not drainages, explained the most genetic variation congruent with results from Funk et al. (2005). Based on radio telemetry, *R. luteiventris* will preferentially travel along stream corridors when available but is capable of traveling through open, dry habitats (Pilliod et al. 2002). These results from both the genetic and movement data suggest topographic barriers limit connectivity among wetlands more than availability of stream corridors for dispersal.

We found a high level of genetic differentiation in *R. luteiventris* consistent with genetic variation generally observed in amphibians (Arens et al. 2007; Shaffer et al. 2000). We identified two primary genetic clusters in STRUCTURE, each with additional subclustering. The northern and southern clusters were highly separable (average q-values by site > 0.85), likely driven by a major ridgeline separating the regions. Identifying substructure within these initial clusters gave a more complete picture of the pattern of genetic variation. Subclusters in the northern region correspond to basins while the first level of substructuring in the south separates Tip Top basin from the remaining three basins (Fig. 12), similar to results found by Funk et al. (2005). Detection of multiple sites with significant bottlenecks in the north suggests the

presence of demographic sinks, likely due to fish introductions (Pilliod and Peterson 2001). This is supported by ongoing monitoring which has not observed a major recruitment event in the northern basins over the past 10 years (D. Pilliod, unpublished data)

Landscape genetics models

In studying metapopulations, it is important to separate and quantify the effects of at site and between site conditions on connectivity (Hayes and Cronin 2004; Gonzalez et al. 1998). In addition, including at site measurements in metapopulation dynamics is vital for assessing population persistence and extinction risk (Baguette and Schtickzelle 2003; Lopez and Pfister 2001). Using gravity models, we were able to separate these effects by demonstrating that distance between sites, at site landscape processes and between site landscape processes are all correlated with connectivity (Table 10).

Isolation-by-distance is common in anuran species (Arens et al. 2007; Beebee 2005; Manier and Arnold 2006), and the importance of distance in explaining connectivity is reflected in all of the selected gravity models. Our observed isolation-by-distance supports the isolation assumption in metapopulation dynamics (Pellet et al. 2007). That is, geographically isolated sites are less connected to other sites in the metapopulation, in contrast to structure observed resulting from general population decline in fragmented systems where populations are equally disjoined (Marsh and Trenham 2001).

At site characteristics are important drivers of metapopulation connectivity. Connectivity generated from a site (i.e., gravity "mass") will be greater when many offspring are produced compared to when few offspring are produced at a site. Landscape processes driving this production of offspring generate the observed source-sink dynamics. In our system, trout are negatively associated with breeding success leading to negligible recruitment and therefore

population sinks (Pilliod et al. 2002). Site depth, as an indirect measure of fish presence, was selected in the saturated unconstrained gravity model. However, in the constrained models the impact of fish is accounted for by the local effect incorporated in the constant term. At site characteristics also identify variables associated with source populations. Larval growth and development is directly related to water temperature (Pough et al. 2004). With the short frost free period in the Bighorn Crags, warming water or melting ice by solar radiation may decrease time to metamorphosis and increase site productivity (Beck and Congdon 2000).

The area of landscape influencing connectivity between sites gives insight into species' ecology. In *R. luteiventris,* bandwidth used to infer landscape influence between sites had little effect. There are two possible explanations for this observed insensitivity to a broader landscape context between sites. In the Bighorn Crags, individuals appear to move via the most direct route (Pilliod et al. 2002). In addition, there is minimal landscape heterogeneity between sites for our selected predictor variables. Therefore, 30m connections captured both the spatial variation between sites and species movement behavior. However, we would expect surrounding landscape context between sites to have more of an effect in fragmented habitats or for species that move by less direct routes.

Elevation is negatively correlated with both genetic diversity (Funk et al. 2005) and gene flow in amphibians (Giordano et al. 2007; Spear et al. 2005). However, elevation or difference in elevation between sites was not selected in any of our models. Although elevation is significantly correlated with restricted gene flow in the above cases, inaccessibility due to high elevation may not be the mechanism. Similar patterns have been observed along latitudinal gradients (Palo et al. 2003), suggesting temperature or length of growing season may be the underlying landscape process. Supporting this argument, temperature was selected in all our

final gravity models, being the first resistance variable retained in all cases. However, all of our sites are at high elevation (>2000m) and have low genetic diversity. Alternately at this fine-scale we may lack the variation in elevation and genetic diversity to see a broader effect.

Ridges are a common barrier for anuran species (Lougheed et al. 1999), demonstrated by high levels of substructure in mountainous areas (Funk et al. 2005; Monsen and Blouin 2003). Although we found evidence for genetic structuring by basin in both the AMOVA and STRUCTURE analyses, ridges were not significant in any of the gravity models. However, topography (elevation relief ratio at a 27X27 cell window size) was highly significant. Elevation relief ratio is a measure of rugosity, indicating high positive or negative dissection of the landscape. At this scale, elevation relief ratio measures ridges, canyons and other major topographic features.

Different landscape processes may govern gene flow across a species range. Our study and Funk et al. (2005) estimate gene flow in *R. luteiventris* at different scales, and provide complimentary information about the species. Broad-scale studies tend to generalize local processes into global averages, elucidating trends that may be unobservable at fine-scales. They give a broader understanding, but may be unable to explain local observations such as metapopulation dynamics. In contrast, fine-scale studies give depth of information about specific landscape processes which may be unique to the study system (Scribner et al. 2001; Spear et al. 2005).

CONCLUSIONS

Gravity models are well-suited for modeling metapopulation connectivity in systems with patches of breeding habitat and an intervening landscape matrix. Measuring connectivity within

a metapopulation is crucial because it generally has a positive effect on population persistence which can be incorporated when estimating thresholds of connectivity (Bowne and Bowers 2004; With 2002). In combination with bottleneck tests (Appendix 12), we were able to identify potential sink populations, an element not incorporated in other approaches. In addition, gravity models are not inherently limited to genetic data. If demographic data are available, estimates of population size could be directly included as an at site (production) parameter. Based on this property, gravity models have the potential to combine both demographic and genetic data in a single connectivity estimate. Finally, once parameters are estimated, gravity models can be used to predict connectivity under alternative landscape conditions such as development alternatives or climate change.

ACKNOWLEDGMENTS

Research was permitted by Idaho Fish and Game (Wildlife Collecting Permit # 030716) and Washington State University Animal Care and Use Protocol (ASAF #3378). R Klaver, P Bartelt, and C. Peterson provided vital project support and feedback. J Evans provided helpful conversations on network applications and R code. A. Robinson provided statistical advice. A Gallant and J. Shive assisted in compiling spatial data. Members of the Palouse Landscape Genetics group provided valuable feedback as the project developed. We would like thank S. Blaser, J. Boone, D. Esperas, E. Letterman, E. Suronen, M. E. Valasquez, and J. Visciglia for field assistance. This work is part of a larger interdisciplinary project on *R. luteiventris* connectivity in the Bighorn Crags funded by USGS – Amphibian Research and Monitoring Initiative (ARMI). Additional support for MAM was provided by EPA-STAR fellowship (FP-916695**)**, NSF-DDIG (grant DEB-0608458 to AS and MAM**)**, and the James King fellowship.

APPENDICES

On-line Appendix 10

Appendix 10 contains general genetic supplementary methods, basic genetic summary statistics, AMOVA tables, and non-significant pair-wise F_{ST} values.

Genetic summary information. To ensure genotype accuracy, we included at least 2 negative controls per extraction and PCR, amplified a known genotype in each reaction, and reamplified all rare alleles (<5% frequency) in addition to at least 10% of samples from each PCR to screen for genotyping and human error. After subsampling sibling clusters (Wang 2004), all loci were tested for global concordance with Hardy-Weinberg equilibrium, significant linkage disequilibrium between all pairs of loci, and Hardy-Weinberg equilibrium at each site using GENEPOP and MICROSATELLITE ANALYSER (MSA) (Dieringer and Schlötterer 2003; Raymond and Rousset 1995). No loci or populations were out of Hardy-Weinberg equilibrium or linkage equilibrium (Raymond and Rousset 1995) with the exception of sites with < 5 samples (see Table 8).

Table A10.1. Microsatellite conditions and basic statistics. Table A10.1 includes the following: locus (Funk et al. 2005; Monsen and Blouin 2003), multiplex (what primers were in a single PCR reaction), primer concentration in PCR reaction, number of alleles found, and expected heterozygosity. Qiagen multiplex kit (including Q-solution) was used for all reactions with standard conditions. General PCR cycle was: initial denaturing (95° C) 15 min, X cycles (94° C 30 sec denature, Y° C annealing 90 sec, 72° C extension 60 sec), and 60° C 60 min final extension.

Locus	Multiplex	[Primer]	Cycles	Annealing	Alleles	H_{exp}
RP17	1	0.31 μM	35	57 °C	3	0.15
RP193	1	0.28 μM	35	57 °C	6	0.52
SRC128	1	0.31 μM	35	57 °C	5	0.08
SFC134	1	0.28 μM	35	57 °C	4	0.55
SFC139	1	0.28 μM	35	57 °C	10	0.76
RP15	2	0.28 μM	40	50.5 °C	3	0.03
RP23	2	0.39 μM	40	50.5 °C	5	0.53
RP3	2	0.28 μM	40	50.5 °C	8	0.72
Average					**5.5**	**0.42**

Table A10.2 **Mantel test** Genetic distance to geographic distance.

Samples	439
Permutations	999

SSx	SSy	SPxy	Rxy	Prob
684683002484.44	2083282.13	238641448.27	0.200	<0.001

Figure A10.1. Non-significant pair-wise F_{ST} values. Black dots represent sites while yellow lines represent non-significant pair-wise comparisons.

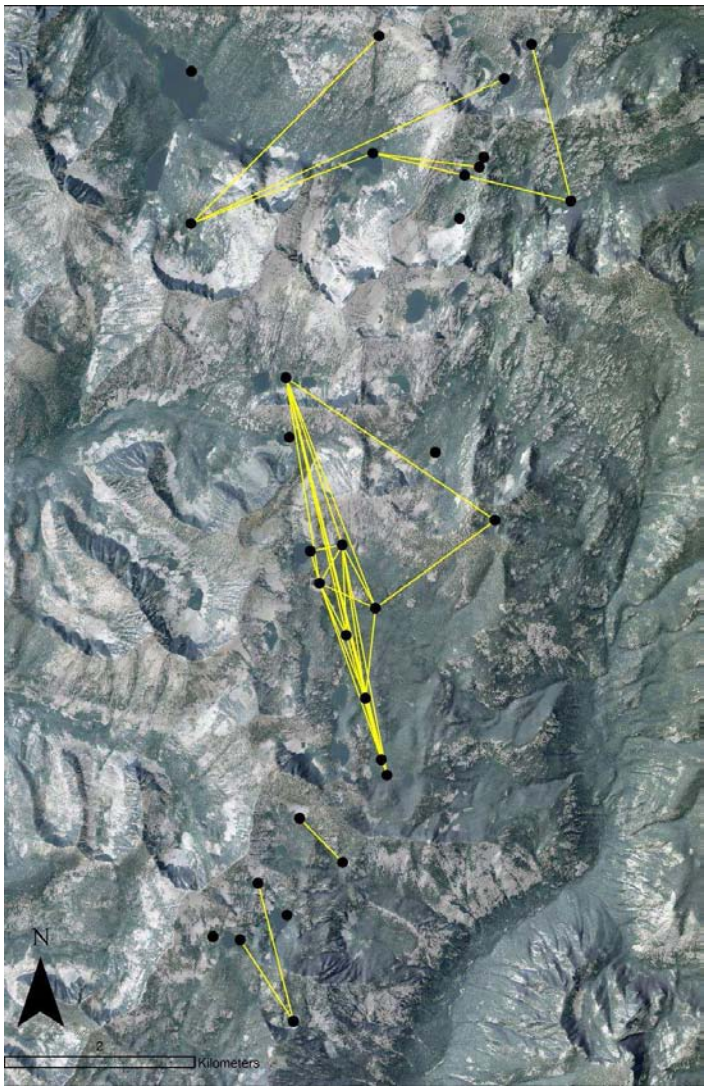
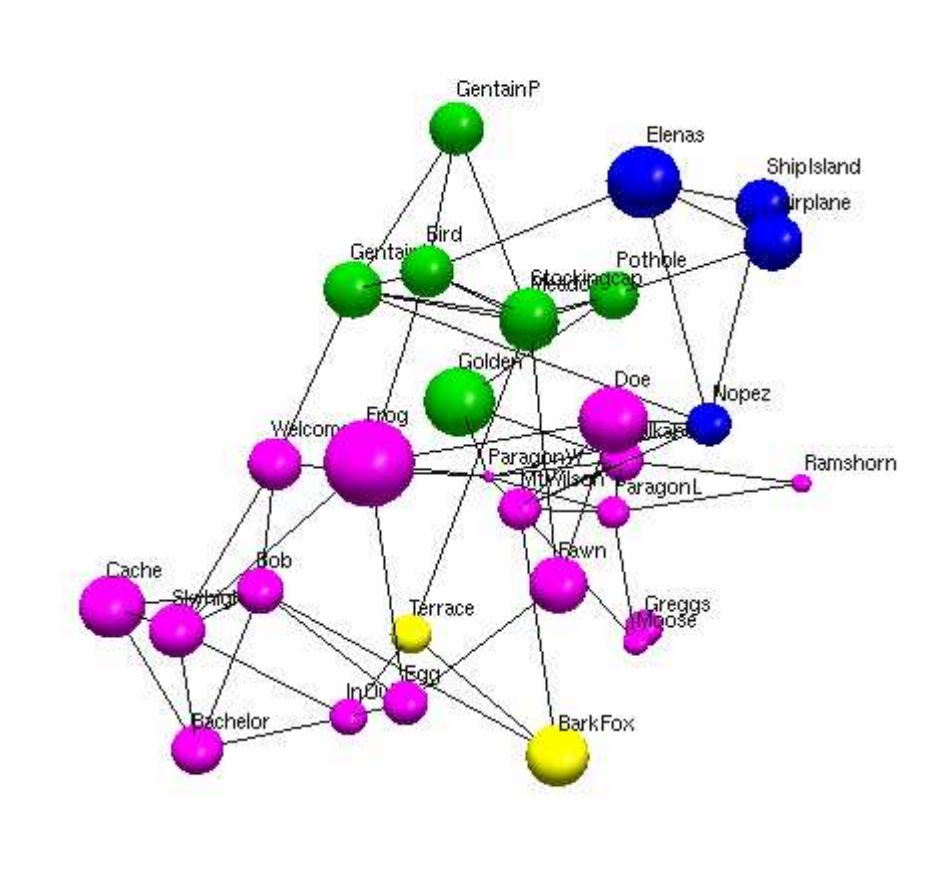


Fig. A10.2. Estimate of most likely connections based on genetic distance (PopGraph). The sites are sized by genetic diversity, colored by drainage (Table 8), and joined by most likely connections based on a maximum likelihood estimate.



On-line Appendix 11

Gravity model justification and derivation

We are presented with the problem of estimating spatial interaction among sample sites (consisting of individuals or groups of individuals), as a function of the landscape. Sites share a proportion of co-dominant alleles from neutral loci such that the spatial distribution of the genetic distances can be estimated by the role of intervening geography. The collection of sites and connections among sites forms a network consisting of nodes (sites) and edges (pair-wise comparisons between sites). Our goal is to estimate "flow" through a network as measured by shared genetic information ("genetic distance"). To do this, we perform the equivalent of estimating a trip distribution (Anderson 1979) from gravity (spatial interaction) models to allocate flows given network of sites:

$$T_{ij} = k v_i^{\mu} w_j^{\alpha} c_{ij}^{-\beta} \quad \text{(eq. 1)}$$

Where T_{ij} is the number of trips between sites, v_i are production/attraction characteristics at the site, w_j is distance to site *j* (or weight given as a function of distance) and c_{ij} is the resistance between locations (Fotheringham and O'Kelly 1989).

We chose to employ a maximum entropy form to justify use of gravity equations for landscape genetic applications. This has several advantages. First, an entropy maximizing approach serves to minimize or eliminate unlikely scenarios ("microstates") resulting in maximized probability (Wilson and Bennett 1985). In addition, the stochastic nature of processes generating genetic variation at neutral loci, mutation and drift, fit well in an entropy context (Brooks and Wiley 1988). Finally, an entropy approach is based on random process and therefore circumvents the need to model individual motivation for movement (Anderson 1979).

The maximum entropy approach permits description of spatial interaction across scales of measurement. Globally ("macrostate"), the system can be described by the flow of genetic information among major regions of the network, which could be identified by genetic clusters. Locally, each sample (individual or group of individuals) is assigned to a single {i, j} position in the origin-destination data matrix. In this approach, instead of trips, we will consider the flow of genetic material as an interaction rate between sites (i.e., migration). The flow of genetic information could be measured by a variety of methods including F_{st} (Wright 1951), D_{ps} - proportion of shared alleles (Bowcock et al. 1994), Nei's genetic distance (Nei 1972), and genetic chord distance (Cavelli-Sforza and Edwards 1967). However, shared alleles and allele frequencies has an intuitive translation in the transportation framework. Therefore, we use D_{ps} in the following justification.

The data matrix takes the form:

		destinations D_j				
		1	2	….	N-1	N
	1	T_{11}	T_{12}	….	$T_{1,N-1}$	T_{1N}
	2	T_{21}	T_{22}	….	$T_{2,N-1}$	T_{2N}
origins O_i	,,,	…	…	….	…	…
	N-1	$T_{N-1,1}$	…	….	…	$T_{N-1,N}$
	N	T_{N1}	T_{N2}	….	$T_{N,N-1}$	T_{NN}

where i and j are row and column indexes of sites. A modified gravity model notation will be employed to describe the spatial genetic context. To be consistent with the definition of interaction, an origin and destination set of sites will be identified. These sites will be termed O_i and D_j corresponding to rows (i) and columns (j) in the spatial interaction data matrix. The entries in the matrix T_{ij} will correspond to the number of "trips" (geneflow) between site i and j and will be represented by D_{ps}. Then $F_1(T_{ij})$ will be the number of microstates giving rise to the

matrix $\{T_{ij}\}$. Note that $T = \sum_i O_i = \sum_j D_j$ is the relative flow of genetic material. The number of ways of selecting n individuals (sample) from N (total population) is $N![n!(N-n)!]^{-1}$ which is set to ${}^{N}B_{n}$. Then combinatorially, the number of ways of selecting and counting all microstates in $\{T_{ij}\}$ is $F_1(\{T_{ij}\}) = T![(\prod_{ij} T_{ij}!)^{-1}]$.

Now that the number of microstates (local effect given by connections) is known, this needs to be maximized to identify the likelihood of the macrostate (global effect). In practice, the function $H = \log F_1(\{T_{ij}\})$ is maximized. The objective function used is $\max H = \log (T! (\prod_{ij} T_{ij}!)^{-1})$ over all elements of $\{T_{ij}\}$ and subject to the interaction constraints $\sum_j T_{ij} = O_i$, $\sum_i T_{ij} = D_j$, $\sum_i \sum_j T_{ij} k_{ij} = K$. But this is a non-linear optimization problem and although usually solved through Lagrange multiplier techniques (Wilson and Bennett 1985), Stirling's approximation provides a simplified objective function by linearizing factorial terms: $H = \log T! - \sum_i \sum_j \log T_{ij}! = \log T! - \sum_i \sum_j (T_{ij} \log T_{ij} – T_{ij})$. The optimization can now be executed using $\partial H/\partial T_{ij} = -\log T_{ij}$ as the objective function. The gravity formulation follows directly arriving at eq. 1 (Fotheringham and O'Kelly 1989; Wilson and Bennett 1985).

For the purposes of landscape genetics, if the problem is reformulated by defining $P_{ij} = T_{ij}/T$ where T_{ij} is a measure of gene flow between i and j then P_{ij} is a probability of an individual case being in the (i,j) state. Then, H can be defined directly as the entropy of the probability distribution $G(P_{ij} \mid \text{for all i,j})$ as $H = -\sum_i \sum_j P_{ij} \log P_{ij}$. This measure then becomes identical with the Shannon information index (Shannon 1948). Maximizing entropy maximizes the likelihood of spatial interaction and expands possible state space.

Gravity models assume a distance effect (w). Isolation-by-distance is commonly seen, and generally expected in the presence of genetic structure (Nei 1972; Nei 1973). Unlike other available landscape genetic approaches, gravity models include both at site characteristics

(production/attraction - v) and between site characteristics (resistance - c). Although landscape condition may restrict migration between sites, at site landscape characteristics may be highly influential. For example, high quality habitat produces more offspring, therefore more opportunities for gene flow from that site. In addition, poor quality habitat may result in breeding failure and therefore no probability of gene flow.

On-line Appendix 12

Appendix 12.1 is an expanded version of Table 7 with metric explanation, source data, predicted effect, and ecological justification. Appendix 12.2 is an expanded version of Table 8 including bottleneck results.

Table of ecological processes and respective independent variables. Parameter – parameter being estimated in the gravity equation. Table headings are as follows. Process – process of which the metric is a measure: isolation by distance (IBD), productivity, predation, habitat permeability and cover (Habitat), topographic morphology (Topo), or temperature-moisture (Temp-Moist). Variable – independent variable name. Code- metric code or abbreviation. Source – source of data which contained the variable or from which the variable was derived. Sources are as follows: NLCD – National Land Cover Database (2001), USGS, SRTM - Shuttle Topographic Radar Mission digital elevation model (DEM), and a climate spline model that adjusts weather stations using elevation (Rehfeldt 2006). Predicted relationship (Pred) – if we expected to find a positive (+), negative (-), or neutral (=) relationship between the independent variable and gene flow. Metric explanation – brief description of variable. Calculation – description of how the metric was calculated or reference. Ecological justification – brief justification for including the variable in our analyses, including selected reference(s).

Some metrics were calculated at multiple window [1]Calcualted at 3X3, 15X15, and 27X27 cell window sizes. [2]Decomposed from 0 – 27X27 cell window size

Parameter	Process	Variable	Code	Source	Pred	Metric Explanation	Calculation	Ecological Justification
Distance	Isolation by distance	Topographic Distance (m)	Distance	SRTM	-	Topographically corrected distance	Sum	Isolation-by-distance (IBD) has been seen in *Rana sp.* (Arens et al. 2007) and generally have little gene flow among populations (Shaffer et al. 2000). In addition, distance seems to be the major limiting factor in *R. luteiventris* connectivity (Pilliod et al. 2002)
Producing/ Attracting	Productivity	Elevation (m)	elev	SRTM	-	Elevation in meters	Elevation at site from DEM	Regulates breeding phenology and site productivity (Funk et al. 2005; Giordano et al. 2007)
		Heat Load Index	hli	SRTM	+	Cool – hot values measuring solar radiation	(McCune and Keon 2002)	Solar radiation controls water temperature and primary productivity. High water temperature associated with breeding (Pilliod et al. 2002)
		Compound topographic index	cti	SRTM	+	Measure of wetness: flow accumulation by catchment size	(Moore et al. 1993)	Hydroperiod is important for amphibian diversity (Babbitt et al. 2003) and *R. luteiventris* prefers deep lakes without fish (Pilliod et al. 2002).
	Predation	Depth	depth	Field	-	Depth of wetland(m)	Wetland depth (Pilliod et al. 2002)	Bigger, deeper wetlands are more likely to have fish predators (Pilliod and Peterson 2001) and is negatively associated with breeding (Pilliod et al. 2002; Watson and McAllister 2003).
Resistance	Habitat	Water	water	NLCD	+	Class Open water	Percent cells of habitat type	Water may provide stopping points during migration (Pilliod et al. 2002)
		Forest	forest	NLCD	+	Classes Deciduous, Evergreen, Mix	Percent cells of habitat type	Forest provides cover (Eigenbrod et al. 2008) and is associated with breeding sites (Pilliod et al. 2002)
		Meadow	Meadow	NLCD	+	Classes: 71, 81, 82	Percent cells of habitat type	Moist areas good for dispersal or seasonal use, importance of vegetative cover (Munger et al. 1998)
	Topo	Elevation relief ratio[1]	err	SRTM	-	Index of elevational complexity	(Evans 1972)	Fine scale – topographic complexity may make travel energetically expensive. Coarse scale – identify major topography which may be barriers (Funk et al. 2005).
		Relative slope position	rsp	SRTM	+	Relative position between valley floor and ridge top	(Murphy et al. submitted)	May be adaptation along attitudinal clines, areas of similar relative position in drainage similar breeding phenology (Bonin et al. 2006; Funk et al. 2005; Giordano et al. 2007)
		Hierarchical slope position[3]	hsp	SRTM	+	Scale decomposition of slope position	(Murphy et al. submitted)	Areas high in the drainage may have reduced genetic variation (Funk et al. 2005).
		Ridges	ridge	SRTM	-	Percent ridge cells between sites	(Murphy et al. submitted)	May be a barrier or semi-permeable barrier to gene flow (Funk et al. 2005; Funk et al. 1999; Lougheed et al. 1999)
	Temp-Moist	Heat Load Index	hli	SRTM	-	Cool – hot values, measure of solar radiation	(McCune and Keon 2002)	Amphibians have little physiological control over water loss, hot areas impede dispersal (Bartelt and Peterson 2005; Duelman and Trueb 1994)
		Compound topographic index	cti	SRTM	+	flow accumulation by catchment size	(Moore et al. 1993)	Wetness may enhance dispersal (Bartelt and Peterson 2005; Pilliod et al. 2002).
		Precipitation ratio	pratio	Spline	+	Growing season precipitation/annual precipitation	(Rehfeldt 2006)	Year-round precipitation, including snow pack, may influencing breeding timing and dispersal (Corn 2003)
		Frost free period	ffp	Spline	+	Julian date of last freeze to date of first freeze	(Rehfeldt 2006)	Later last freeze means a shortened breeding season, which may result in restricted gene flow (Palo et al. 2003).

Drainage	Basin	Site	Cluster	Genotypes	Basin Total	Drainage Total	Sign	Wilcoxon	Mode shift
Clear	Birdbill	Birdbill Lake[1]	IA	4			na	na	na
		Gentian Lake	IA	20			n,n	**ex,ex**	**y**
		Gentian Ponds	IA	18			n,n	n,n	n
		Meadow Lake	IA	20			n,n	n,n	n
		Mirror Lake[1]	I*	1			na	na	na
		Stocking Cap Lake	IA	6	69		n,n	**ex,ex**	**y**
	Glacier	Golden Lake	IB	6			n,n	n,n	y
		Pothole Lake	IB	13	19	88	n,n	**ex,ex**	**y**
Nopez	Nopez	Nopez Lake	IC	5			n,n	n,n	y
		Ship Island Lake	IC	18	23		y,n	**ex,ex**	**y**
	Sheepeater	Airplane Lake	IC	22			n,n	n,def	N
		Elenas Lake	IC	20	42	65	n,n	n,n	N
Waterfall	Terrace	Barking Fox Lake	IIB	14			n,n	n,n	N
		Terrace Lakes[2]	IIA	10	24	24	n,n	n,n	Y
Wilson	Harbor	Bob Lake	IIA	17			n,n	n,n	N
		Buteo Meadow[1]	II*	4			na	na	Na
		Tobias Lake[1]	IIA	3			na	na	Na
		Welcome Lake	IIA	15	39		n,y	n,def	N
	Skyhigh	Bachelor Meadow	IIA	8			n,n	n,n	Y
		Buck Lake[1]	IIA	3			na	na	Na
		Cache Lake	IIA	7			n,n	**ex,ex**	**Y**
		Doe Lake	II*	6			n,n	n,n	Y
		Egg White Lake	IIA	17			n,n	ex,ex	Y
		Fawn Lake	IIA	17			n,n	n,n	N
		Frog Pond Lake	IIA	10			n,n	n,n	N
		Glacial Lake[1]	IIA	2			na	na	Na
		Homer Pond[1]	II*	1			na	na	Na
		In and Out Lake	IIA	10			n,n	n,n	N
		Skyhigh Lake	IIA	23			n,n	def,n	N
		Mount Wilson	IIA	15			n,n	n,n	N
		Twin Cove Lake	IIA	7	126		n,n	n,n	Y

Drainage	Basin	Site	Cluster	Genotypes	Basin Total	Drainage Total	Sign	Wilcoxon	Mode shift
	Tip Top	Greggs Lake	IIB	20			n,n	n,n	N
		Moose Lake	IIB	18			n,n	n,n	N
		Paragon Lake	IIB	20			n,n	def,n	N
		Ramshorn Lake	IIB	18			n,n	n,n	N
		U. Paragon	IIB				n,n	n,n	Y
		Wetland		6					
		Walkabout Lake	IIB	17	99	264	n,n	def,n	N
Total				441					

LITERATURE CITED

Aikaike, H. 1973, Information theory as an extension of the maximum likelihood principle. B. N. Petrov, and F. Csaki, eds. Second International Symposium on Information Theory.

Anderson, J. E. 1979. A theoretical foundation for the gravity equation. American Economic Review 69:106-116.

Arens, P., T. van der Sluis, W. P. van't Westende, B. Vosman, C. C. Vos, and M. J. M. Smulders. 2007. Genetic population differentiation and connectivity among fragmented Moor frog (*Rana arvalis*) populations in The Netherlands. Landscape Ecology 22:1489-1500.

Babbitt, K., M. Baber, and T. Tarr. 2003. Patterns of larval amphibian distribution along a wetland hydroperiod gradient. Canadian Journal of Zoology 81:1539-1552.

Baguette, M., and N. Schtickzelle. 2003. Local population dynamics are important to the conservation of metapopulations in highly fragmented landscapes. Journal of Applied Ecology 40:404-412.

Bartelt, P., and C. R. Peterson. 2005. Physical modeling operative temperature and evaporative water loss in amphibians. Journal of Thermal Biology 30:93-102.

Beck, C. W., and J. D. Congdon. 2000. Effects of age and size at metamorphosis in *Hyla squirella*. Oecologia 112:187-192.

Beebee, T. J. C. 2005. Conservation genetics of amphibians 95:423-427.

Bonin, A., P. Taberlet, C. Miaud, and F. Pompanon. 2006. Explorative Genome Scan to Detect Candidate Loci for Adaptation Along a Gradient of Altitude in the Common Frog (*Rana temporaria*). Mol Biol Evol 23:773-783.

Bosch, J., P. A. RinconI, L. Boyero, and I. Martinez-Solano. 2006. Effects of introduced salmonids on a montane population of Iberian frogs. Conservation Biology 20:180-189.

Bossenbroek, J., L. E. Johnson, B. Peters, and D. M. Lodge. 2007. Forecasting the expansion of zebra mussels in the United States. Conservation Biology 21:800-810.

Bossenbroek, J., C. E. Kraft, and J. C. Nekola. 2001. Prediction of long-distance dispersal using gravity models: Zebra mussle invasion of inland lakes. Ecological Applications 11:1778-1788.

Bouchy, P., K. Theodorou, and D. Couvet. 2005. Metapopulation viability: influence of migration. Conservation Genetics 6:75-85.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455-457.

Bowne, D. R., and M. A. Bowers. 2004. Interpatch movements in spatially structured populations: a literature review. Landscape Ecology 19:1-20.

Brooks, D. R., and E. O. Wiley. 1988, Evolution as entropy: toward a unified theory. Chicago, Il, University of Chicago Press.

Castellano, S., and E. Balletto. 2002. Is the partial mantel test inadequate? Evolution 56:1871-1873.

Cavelli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. American Journal of Human Genetics 19:233-257.

Collins, J. P., and A. Storfer. 2003. Global amphibians declines: sorting the hypotheses. Diversity and Distribution 9:89-98.

Compton, B. W., K. McGarigal, S. A. Cushman, and L. R. Gamble. 2007. A resistant-kernel model of connectivity for amphibians that breed in vernal pools. Conservation Biology 21:788-799.

Corn, P. S. 2003. Amphibian breeding and climate change: importance of snow in the mountains. Conservation Biology 17:622-625.

Cornuet, J. M., and G. Luikart. 1997. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001-2014.

Dieringer, D., and C. Schlötterer. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3:167-169.

Duelman, W., and L. Trueb. 1994, The biology of amphibians. Baltimore, Maryland, Johns Hopkins University Press.

Dyer, R. J., and J. D. Nason. 2004. Population graphs: the graph theoretic shape of genetic structure. Molecular Ecology 13:1713-1727.

Eigenbrod, F., S. J. Hecnar, and L. Fahrig. 2008. The relative effects of road traffic and forest cover on anuran populations. Biological Conservation 141:35-46.

Evans, I. S. 1972. General geomorphometry, detivatives of altitude, and descriptive statistics *in* R. J. Chorley, ed. Spatial Analysis in geomorphology. New York, Harper & Row.

Ferrari, M. J., O. N. Bjornstad, J. L. Partain, and J. Antonovics. 2006. A gravity model for the spread of a pollinator-borne plant pathogen. American Naturalist 168:294-202.

Fotheringham, A. S., and M. E. O'Kelly. 1989, Spatial interaction models: formulation and applications. Dordrecht, Kluwer Academic.

Funk, W. C., M. S. Blouin, P. S. Corn, B. A. Maxell, D. S. Pilliod, S. Amish, and F. W. Allendorf. 2005. Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. Molecular Ecology 14:483-496.

Funk, W. C., D. A. Tallmon, and F. W. Allendorf. 1999. Small effective population size in the long-toed salamander. Molecular Ecology 8:1633-1640.

Gardner, T. A., J. Barlow, and C. A. Peres. 2007. Paradox, presumption and pitfalls in conservation biology: The importance of habitat change for amphibians and reptiles. Biological Conservation 138:166-179.

Gilpin, M. 1991. The genetic effective size of a metapopulation. Biological Journal of the Linnean Society 42:165-175.

Giordano, A. R., B. J. Ridenhour, and A. Storfer. 2007. The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactylum*). Molecular Ecology 16:1625-1637.

Goldberg, C. S., M. E. Kaplan, and C. R. Schwable. 2003. From the frog's mouth: buccal swabs for collection of DNA from amphibians. Herpetological Review 34:220-221.

Gonzalez, A., H. J. Lawton, F. S. GIlbert, T. M. Blackburn, and I. Evans-Freke. 1998. Metapopulation dynamics, abundance, and distribution in a microecosystem. Science 281:2045-2047.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). jerome.goudet@ie-zea.unil

Green, D. M. 2003. The ecology of extinction: population fluctuation and decline in amphibians. Biological Conservation 111:331-343.

Hanski, I., and O. E. Gaggiotti. 2004. Ecology, genetics, and evolution of metapopulations, Pages 696 *in* C. Crumly, ed. Burlington MA, Elsevie

Hayes, K. J., and J. T. Cronin. 2004. Confoudning of patch quality and matrix effects in herbivore movement studies. Landscape Ecology 19:119-124.

Jiang, Y., S. K. Swallow, and P. W. C. Paton. 2007. Designing a spatially-explicit nature reserve network based on ecological functions: An integer programming approach. Biological Conservation 140:236-249.

Johnson, J. R., and R. D. Semlitsch. 2003. Defining core habitat of local populations of the gray treefrog (*Hyla versicolor*) based on choice of oviposition site. Oecologia 137:205-210.

Keyghobadi, N., J. Roland, and C. Strobeck. 2005. Genetic differentiation and gene flow among populations of the alpine butterfly, Parnassium smintheus, vary with landscape connectivity. Molecular Ecology 14:1897-1909.

Legendre, P. 2000. Comparison of permutation methods for the partial correlation and partial mantel tests. Journal of Statistical and Computational Simulation 67:37-73.

Lopez, J. E., and C. A. Pfister. 2001. Local population dynamics in metapopulation models: implications for conservation. Conservation Biology 15:1700-1709.

Lougheed, S. C., C. Gascon, D. A. Jones, J. P. Bogard, and P. T. Boag. 1999. Ridges and rivers: a test of competing hypotheses of Amazonian diversification using a dart-poison frog (*Epipedobates femoralis*). Proceedings of the Royal Society B 266:1829-1835.

Luikart, G., and J. Cornuet. 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology 12:228-237.

Luikart, G., J. M. Cornuet, F. W. Allendorf, and W. B. Sherwin. 1998a. Distortion of allele frequency distributions provide a test for recent population bottlenecks. Journal of Heredity 89:238-247.

Luikart, G., W. B. Sherwin, B. M. Steele, and F. W. Allendorf. 1998b. Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. Molecular Ecology 7:963-974.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189-197.

Manier, M. K., and S. J. Arnold. 2006. Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. Proceedings of the Royal Society B 273:3001-3009.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

Marsh, D. M., and P. C. Trenham. 2001. Metapopulation dynamics and amphibian conservation. Conservation Biology 15:40-49.

McCune, B., and D. Keon. 2002. Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science 13:603-606.

McRae, B. H. 2006. Isolation by resistance. Evolution 60:1551-1561.

Monsen, K., and M. S. Blouin. 2003. Genetic structure in a montane ranid frog: restricted gene flow and nuclear-mitochondrial discordance. Molecular Ecology 12:3275-3286.

Moore, I., P. Gessler, G. Nielsen, and G. Petersen. 1993. Terrain attributes and estimation methods and scale effects, Pages 189-214 *in* A. Jakeman, M. Beck, and M. McAleer, eds., Modeling change in environmental systems. London, Wiley.

Munger, J. C., M. Gerber, K. Mardrid, M.-A. Carroll, W. Patersen, and L. Heberger. 1998. U.S. National Wetland Inventory classifications as predictors of the occurrence of Columbia spotted frogs (*Rana luteiventris*) and Pacific treefrogs (*Hyla regilla*). Conservation Biology 12:320-330.

Murphy, M. A., J. S. Evans, and A. Storfer. in review. Quantifying ecological process at multiple spatial scales using landscape genetics: *Bufo boreas* connectivity in Yellowstone National Park. Ecology.

Nei, M. 1972. Genetic distance between populations. The American Naturalist 106:283-292.

—. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Science 70:3321-3323.

Orizaola, G., and F. Brana. 2006. Effect of salmonid introduction and other environmental characteristics on amphibian distribution and abundance in mountain lakes of northern Spain. Animal Conservation 9:171-178.

Ovaskainen, O., and I. Hanski. 2004. Metapopulation dynamics in highly fragmented landscapes, Pages 73-104 *in* I. Hanski, and O. E. Gaggiotti, eds. Ecology, genetics, and evolution of metapopulations. Burlington MA, Elsevier.

Palo, J. U., R. B. O'Hara, A. T. Laugen, A. Laurila, C. R. Primmer, and J. Merila. 2003. Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural

selection: evidence from a comparison of molecular and quantitative genetic data. Molecular Ecology 12:1963–1978.

Peakall, R., and P. E. Smouse. 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Peakall, R., P. E. Smouse, and D. Huff. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. Molecular Ecology 4:135-147.

Pechmann, J. H. K., D. E. Scott, R. D. Semlitsch, J. P. Caldwell, L. J. Vitt, and J. W. Gibbons. 1991. Declining amphibian populations: the problem of separating human impacts from natural fluctuations. Science 253:892-895.

Pellet, J., E. Fleishman, D. S. Dobkin, A. Gander, and D. D. Murphy. 2007. An empirical evaluation of the area and isolation paradigm of metapopulation dynamics. Biological Conservation 136:483-495.

Pilliod, D. S., and C. R. Peterson. 2001. Local and landscape effects of introduced trout on amphibians in historically fishless watersheds. Ecosystems 4:322-333.

Pilliod, D. S., C. R. Peterson, and P. I. Ritson. 2002. Seasonal migration of Columbia spotted frogs (*Rana luteiventris*) among complementary resources in a high mountain basin. Canadian Journal of Zoology 80:1849–1862.

Pough, F. H., R. M. Andrews, J. E. Cadle, M. L. Crump, A. H. Savitzky, and K. D. Wells. 2004, Herpetology. Upper Saddle River, Pearson Prentice Hall.

Prichard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pulliam, H. R. 1988. Sources, sinks and population regulation. American Naturalist 132:652-661.

R Development Core Team. 2007. R: A language and environment for statistical computing, version 2.7.0. R Foundation for Statistical Computing, Vienna, Austria.

Raufaste, N., and F. Rousset. 2001. Are partial mantel tests adequate? Evolution 55:1703-1705.

Raymond, M., and F. Rousset. 1995. GENEPOP(1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248-249.

Rehfeldt, G. E. 2006. A spline model of climate for the Western United States, 21 p, General Technical Report RMRS-GTR-165. Fort Collins, CO, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Ricketts, T. 2001. The matrix matters: effective isolation in fragmented landscapes. American Naturalist 158:87-99.

Rousset, F. 2002. Partial mantel tests: reply to Castellano and Balletto. Evolution 56:1874-1875.

Scribner, K., J. W. Arntzen, N. Cruddace, R. S. Oldham, and T. Burke. 2001. Environmental correlates of toad abundance and population genetic diversity. Biological Conservation 98:201-210.

Shaffer, H. A., G. M. Fellers, A. Magee, and S. R. Voss. 2000. The genetics of amphibian declines: population substructure and molecular differentiations in the Yosemite toad, *Bufo canorus* (Anura, Bufonidae) based on single-strand conformation polymorphism analysis (SSCP) and mitochondrial DNA sequence data. Molecular Ecology 9:245-259.

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal 27:379-423, 623-656.

Smith, A. M., and D. M. Green. 2005. Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations? Ecography 28:110-128.

Spear, S. F., C. R. Peterson, M. Matacq, and A. Storfer. 2005. Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). Molecular Ecology 14:2553-2564.

Storfer, A. 2003. Amphibian declines: future directions. Diversity and distributions 9.

Storfer, A., M. A. Murphy, J. S. Evans, C. Goldberg, S. Spear, S. Robinson, R. Dezzani et al. 2007. Putting the 'landscape' in landscape genetics. Heredity 98:128-142.

Stuart, S. N., J. S. Chanson, N. A. Cox, B. E. Young, A. S. L. Rodrigues, D. L. Fischman, and R. W. Waller. 2004. Status and trends of amphibian declines and extinctions worldwide. Science 306:1783-1786.

Voorhees, A. M. 1956, A genetial theory of traffic movement Institute of traffic engineering.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics 166:1963-1979.

Watson, J., and K. McAllister. 2003. Home ranges, movements, and habitat selection of Oregon spotted frogs (*Rana pretiosa*). Journal of Herpetology 37:292-300.

Wiens, J. A. 1996. Wildlife in patchy environments: Metapopulations mosaics, and management in D. R. McCullough, ed. Metapopulations and Wildlife Conservation. Washington, DC, Island press.

Willig, R. D., and E. E. Bailey. 1979. The economic gravity model. American Economic Review 69:96-101.

Wilson, A. G., and R. J. Bennett. 1985, Mathematical methods in human geography and planning, J. Wiley and Sons, Inc.

With, K. A. 2002. Using percolation theory to assess landscape connectivity and effects of habitat fragmentation in K. J. Gurzwiller, ed. Applying Landscape Ecology in Biological Conservation. New York, Springer-Verlag.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

Xia, Y., N. B. Ottar, and B. T. Grenfell. 2004. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. American Naturalist 164:267-281.