LANDSCAPE GENETICS OF THE AMERICAN BADGER: UNDERSTANDING CHALLENGES IN ELUSIVE SPECIES

by

Elizabeth M Kierepka

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Biological Sciences

at

The University of Wisconsin-Milwaukee

December 2014

ABSTRACT

LANDSCAPE GENETICS OF THE AMERICAN BADGER: UNDERSTANDING CHALLENGES IN ELUSIVE SPECIES

by

Elizabeth M Kierepka

The University of Wisconsin-Milwaukee, 2014

Under the Supervision of Professor Emily Latch

American badgers are one of the most poorly understood carnivores in North America due to their highly elusive, semifossorial lifestyle. To complicate understanding their biology further, badgers possess life history characteristics that predict radically different responses to habitat heterogeneity. In particular, they are considered grassland specialists, so their movement and population viability could be highly dependent on grassland habitats. Badgers are also highly mobile, which suggests they experience high gene flow. Predicting how these life history traits impact gene flow, however, is difficult based on the high diversity in responses to landscape heterogeneity among carnivores. To assess how landscape heterogeneity affects gene flow in badgers, my dissertation contains three chapters. In the first chapter, I assessed performance of individual-based landscape genetic methods to identify statistics that would be most appropriate for elusive species like badgers. Once I identified methods that would be appropriate for badgers, I combined both individual-based simulations and landscape genetic methods in my second chapter to assess how landscape heterogeneity in Wisconsin affects gene flow in a protected population of badgers. My final chapter investigated how historical and

contemporary changes in grasslands has affected genetic variation in badgers across their North American range.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (SUPPLEMENTARY)

LIST OF TABLES

LIST OF TABLES (SUPPLEMENTARY)

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Emily Latch, who has been both an inspiration and mentor throughout my Ph.D. graduate career. Through her encouragement and patience, I owe much of my growth in passion and confidence as a scientist to her. I also would like to extend gratitude to my committee members, Sara Hoot, Linda Whittingham, Peter Dunn, and Filipe Alberto, whose guidance and insights greatly aided in my abilities to think critically as a scientist. I am genuinely honored to have worked with all of my committee members, and thank them for their contributions in my graduate career.

Badgers are highly elusive, and without the help of countless individuals in the state of Wisconsin and across North America, I would have never collected any samples. In particular, I would like to thank Karl Martin, John Olson, Brian Sloss, and Dave Sample for their constant enthusiasm and support for the Wisconsin badger project. Many individuals were involved with sample collection for my range wide analysis, but Brian MacMillan at the North American Fur Auction and Bradley Swanson at Central Michigan University were especially critical for the success of my project. I am grateful to all the Wisconsin Department of Natural Resources employees, citizens of Wisconsin, and statewide wildlife personnel that reported and sent me samples or photos of badgers. Finally, I wanted to thank Katie McKenney and the staff at UW-Madison Biotechnology Center for laboratory assistance.

My funding sources were integral to the completion of this project, particularly funding from the Wisconsin Department of Natural Resources and the University of Wisconsin-Milwaukee. Funding was provided via the WDNR Federal Wildlife Restoration Grant W-160-P) and the University of Wisconsin-Milwaukee Graduate School (Distinguished Dissertation Fellowship) for financial support. The Department of Biological Sciences at the University of Wisconsin-Milwaukee Graduate School (Ruth Walker Award, Ivy Balsam Award, James D. Anthony Award, Joseph Baier Award, and Chancellor's Award) and the American Society of Mammalogists (Grant-in-Aid and Anna Jackson Award) also provided much needed funding for my laboratory and field work.

Family and friends provided much needed support throughout my graduate career, and I cannot express how much I appreciate all their enthusiasm and love. I would fill up another dissertation worth of thank-yous to you all, but I doubt I would have kept my sanity or passion for science with all of your constant smiles, late night chat sessions, and round-table discussions. Most of all, thank you to my family for their constant love and unwavering support.

CHAPTER 1: PERFORMANCE OF PARTIAL STATISTICS IN INDIVIDUAL-BASED LANDSCAPE GENETICS

Introduction

Classical population genetics and contemporary landscape genetics traditionally involve analyzing genetic differentiation within and among genetically discrete groups. Populations can be challenging to delineate in continuously distributed species that do not have clearly defined population boundaries. One approach for delineating cryptic population structure is to utilize Bayesian clustering algorithms (e.g., Pritchard *et al*. 2000; Guillot *et al*. 2005; Corander & Marttinen 2006) to define populations within a landscape. However, common patterns of spatial genetic structure such as isolation-by-distance (IBD; Wright 1943) or weak barriers to gene flow can lead to incorrect estimates of the number of genetically discrete populations across a landscape (Latch *et al*. 2006; Frantz *et al*. 2009; Schwartz & McKelvey 2009). In these situations, individual-based analytical approaches offer a viable alternative to population-based techniques. Because individual-based statistics do not require *a priori* definition of populations, they offer an appealing alternative approach to investigate a wide variety of fine-scale influences on gene flow.

Individual-based methods in landscape genetics have become particularly popular for differentiating between two models of gene flow across a heterogeneous landscape: isolation by distance, and isolation-by-resistance (IBR; McRae 2006). IBD occurs when geographic distance is positively correlated with genetic differentiation regardless of

landscape composition, and often serves as a null hypothesis in landscape genetic studies. Conversely, in IBR, landscape heterogeneity (e.g., habitat types, roads, rivers) influences gene flow. In order to evaluate the relative influence of habitat variables on gene flow, the landscape itself is parameterized using user defined resistance values. These resistance values are assigned to each pixel on a raster map based on their hypothesized effect on gene flow. For example, Cushman *et al*. (2006) parameterized landscapes for black bears in Idaho such that anthropogenic structures like roads and agriculture inhibited gene flow (high resistance) and suitable forest habitats promoted gene flow (low resistance). Selection of appropriate resistance values requires *a priori* knowledge of relevant landscape features (for discussion see Zeller *et al*. 2012), so previous life history data such as telemetry (e.g., Cushman & Lewis 2010; Reding *et al*. 2013), expert opinion (e.g., O'Brien *et al*. 2006), or presence/absence data (e.g., Walpole *et al*. 2012) are often employed to derive meaningful resistance surfaces (raster maps where each pixel has a resistance value). Each parameterized resistance surface, therefore, represents a hypothesis about how landscape features affect gene flow. Hypothesis testing then involves quantifying correlations between landscape resistance and pair-wise measures of genetic differentiation to identify which landscape features, if any, explain patterns of genetic differentiation. One potential challenge to quantifying IBR is that landscape parameters of interest are often tightly correlated with geography. Autocorrelation in landscape and genetic variables can then result in spurious correlations in genetic analyses unless statistics can explicitly incorporate geographic covariates (Meirmans *et al*. 2012).

One of the primary remedies for spatial autocorrelation in IBR studies has been the use of partial statistics, in which the effects of geographic coordinates are separated from landscape data. Partial statistics can isolate IBR from IBD, and therefore, prevent spurious correlations observed in aspatial statistics (e.g., Cushman & Landguth 2010; Meirmans *et al*. 2012). Currently, the most prominent partial statistics in individual-based landscape genetics can be broadly classified as distance-based methods. Distance-based methods typically correlate pair-wise distance metrics of landscape resistance with genetic differentiation (termed genetic distance) while controlling for straight-line (Euclidean) distances between individuals. Numerous metrics exist for estimating genetic differentiation between pairs of individuals, where larger calculated values of genetic distance typically indicate higher levels of divergence. Pair-wise measures of cost to gene flow through a landscape between all pixels of the landscape (hereby called cost distance) are then calculated based on resistance surfaces. Cost distances are essentially the cumulative effect of traveling through pixels of different resistances between two individuals, irrespective of geographic distance, so like resistance values, higher cost distances represent stronger impediments to gene flow. Generally, once pair-wise distance measures are derived, distance-based partial statistics test for positive correlations between genetic and cost distances indicative of IBR. Although several distance-based methods can be adapted for genetic data, the partial Mantel test (Smouse *et al*. 1986) is unquestionably the most popular in landscape genetics.

Mantel tests quantify correlations between two or more matrices, and in landscape genetics, these matrices are typically composed of pair-wise genetic, Euclidean, and cost distances. Based on their widespread use in landscape genetics, simulation studies have

largely focused on identifying factors that affect performance of Mantel tests like resistance strength (i.e., values of pixels within raster maps; Cushman *et al*. 2012) and landscape configuration (Landguth *et al*. 2012). Resistance strength and landscape configuration have received attention in simulation studies because both can dramatically affect conclusions about gene flow (Cushman *et al*. 2012; Oyler-McCance *et al*. 2013). In general, Mantel tests perform best in highly fragmented landscapes with high amounts of resistant matrix habitat (Cushman *et al*. 2012). When resistance values of matrix habitat are much higher than suitable habitat (generally 5 to 10 times higher), partial Mantel tests detect inhibitory effects of habitat variables on gene flow (Cushman *et al*. 2012). Even when landscape configuration and resistance values are set to maximize statistical performance within simulated datasets, multiple authors have raised concerns about potential problems with Mantel tests including low power (Legendre & Fortín 2010; Graves *et al*. 2013) and high type I error rates (Balkenhol *et al*. 2009; Guillot & Rousset 2013). These studies have provided important caveats for empirical studies that rely on Mantel tests, and elucidate the need for alternative statistical approaches to complement results from Mantel tests and improve interpretation of empirical datasets.

Few other partial statistics have received the mass utilization as Mantel tests, but ordination techniques such as redundancy analysis (RDA) have recently been suggested as a viable alternative (e.g., Balkenhol *et al*. 2009; Legendre & Fortín 2010). Ordination techniques offer considerable flexibility as compared to Mantel tests because they do not require distance-based metrics and can overcome core assumptions of Mantel tests (e.g., linear relationships between variables). RDA is an unconstrained ordination technique where genetic data can be either genetic distances (distance-based redundancy analysis

[dbRDA]; Legendre and Anderson 1999) or individualistic measures of genetic variation like allele frequencies. Individualistic metrics for landscape and geographic variables (e.g., spatial coordinates, habitat type identity, precipitation measures, side of putative barrier) increase the diversity of explanatory variables that can be tested while negating the requirement for pairwise distance calculations that can result in losses of statistical power (Legendre & Fortín 2010). Also, ordination techniques like RDA and dbRDA can utilize transformations such as principal coordinates analysis (PCoA) to linearize genetic variables, thus removing any potential violations of linearity observed in Mantel tests (Graves *et al*. 2013).

Another potential advantage of ordination methods is that they provide measures of variance around parameter estimates, which allows for improved interpretation of results over Mantel tests that only provide a correlation coefficient and *p*-value (Legendre & Fortín 2010). Correlation coefficients from Mantel tests can be highly variable (Guillot & Rousset 2013; Graves *et al*. 2013), and without measures of variance, researchers may have trouble distinguishing between a type I error and the correct conclusion. In contrast, RDA provides statistics typical of ANOVAs including sum of squares, variance explained by each component, and measures of variance around F-ratios. These additional results generated by RDA provide more information upon which to base interpretations about whether and how particular landscape variables influence gene flow (i.e., partition variance among landscape variables) and the statistical validity of the entire model (Legendre & Fortín 2010). Given these characteristics – the ability to overcome linearity assumptions, flexibility in type and number of explanatory variables,

and statistical outputs – ordination methods offer potentially useful alternatives to Mantel tests.

Within individual-based genetics, simulation studies have evaluated the performance of Mantel tests under different scenarios that occur in empirical datasets (e.g., Cushman *et al*. 2012, 2013). Our goal was to use simulations to test performance of ordination techniques, a potentially viable alternative to Mantel tests for individual-based landscape genetic analyses. We quantified the performance of partial Mantel tests and ordination methods by assessing each statistic's accuracy and explanatory power for detecting IBR (alternative hypothesis where landscape parameter was significant) while controlling for underlying patterns of IBD (null hypothesis of no effect of landscape heterogeneity). Simulations were designed such that spatial genetic structure was driven by IBD in all landscapes, and additionally included varying levels of resistance across the landscape (IBR). Given that RDA does not require quantification of pair-wise distances and tend to be more robust in population-based analyses (Balkenhol *et al*. 2009; Legendre & Fortin 2010), RDA was predicted to be more effective at detecting the effect of landscape resistance (i.e., costs to gene flow) than Mantel tests. In addition to comparing statistical approaches, we also designed simulations to assess the role of underlying landscape structure (habitat amount and configuration) and resistance strength of matrix habitat on statistical accuracy and explanatory power of IBR models. Both landscape structure and resistance strength have been identified as important predictors of accuracy in Mantel tests (Cushman *et al*. 2012), and we expected that these variables would likewise influence performance in ordination methods, though perhaps to a different

degree. Our findings are relevant to the larger field of work investigating optimal methods for detecting IBR in individual-based landscape genetics.

Methods

Generation of Landscapes

Landscapes in nature are composed of a mosaic of features and resources that each species can perceive as suitable or unsuitable based on their life history requirements. Many species utilize areas within a heterogeneous landscape in a non-random way, and preference for or avoidance of certain landscape variables is often assumed to promote or inhibit gene flow respectively (e.g., Shafer *et al*. 2012). Our simulated landscapes were composed of two habitat types: suitable and unsuitable (matrix) habitats. Heterogeneity of a landscape can be expressed using many different metrics, but we controlled two aspects of landscape configuration, level of fragmentation and proportion of suitable habitat, because these aspects have been shown to impact the strength of correlations between genetic differentiation and landscape resistance (Cushman *et al*. 2012). The level of fragmentation, in particular, seems to have direct impacts on both demographic (Revilla & Wiegand 2008) and genetic processes (Keyghobadi 2007; Bruggeman *et al*. 2010) in empirical and simulated datasets, so controlling fragmentation and proportion of habitat types should allow us to isolate the effects of different cost measures and resistance strength on partial statistics.

We manipulated habitat configuration of the simulated landscapes by creating artificial landscapes in QRULE (Gardner 1999) using the "Multifactal Random Map"

function. Multifactal random maps were created by defining the amount of aggregation of pixels or level of fragmentation (H) as well as the proportion of suitable habitat (P). We chose an H value of 0.50 with suitable habitat comprising either 20%, 40%, 60%, or 80% of the total landscape because extreme values in H or P in either direction are probably less realistic in nature or prevent meaningful levels of IBR or IBD within simulations. Only variation in P was tested within fractal landscapes because Cushman *et al*. (2012) found that configuration metrics (e.g., patch cohesion; Schumaker 1996; clumpy index of habitat aggregation; McGarigal *et al*. 2002, and correlation length of habitat; McGarigal *et al*. 2002) that explained genetic differentiation were highly correlated with P, but not with H. Therefore, varying P is likely a more effective way to create landscapes that result in detectable levels of genetic differentiation. Five replicate landscapes were generated within each P category (20%, 40%, 60%, or 80% suitable habitat) to evaluate variation in statistics, resulting in 20 independent landscape configurations (Figure 1.1).

We calculated three configuration metrics (patch cohesion, clumpy index of habitat aggregation, and correlation length of habitat) for all simulated landscapes. Configuration metrics have higher predictive power to explain how landscapes impact IBR than just suitable habitat alone (Cushman *et al*. 2012). Patch cohesion, a measure of habitat class aggregation, measures the physical connectedness of a focal habitat class as a percentage (McGarigal *et al*. 2002). Clumpy index of habitat aggregation (hereby called clumpy) also measures habitat class aggregation, but is not influenced by area of a focal habitat class (Neel *et al*. 2004). Clumpy ranges from -1 to 1 where -1 is completely disaggregated, 0 is random, and 1 is maximally clumped (McGarigal *et al*. 2002).

Finally, correlation length of habitat defines the average distance an individual could travel in a random direction without leaving suitable habitat when randomly placed in a habitat patch, making correlation length a measure of the extensiveness of suitable habitat (McGarigal *et al*. 2002).

By utilizing multiple landscapes from QRULE, we were able to isolate the impact of different resistance strengths in addition to suitable habitat amount and configuration on detection of IBR. Each simulated raster was 512 X 512 pixels (each pixel = 10 m x 10 m) where each pixel contained one of two cost values: suitable habitat was always set to 1 and the matrix was given a value of 1 (indicating IBD only), 5, 10, 25, or 50. Essentially, matrix habitat was parameterized such that high valued pixels impeded gene flow, and the higher the value, the stronger the inhibitory effect. In total, five landscapes (R1, R5, R10, R25, and R50) per habitat area category (total = 100 landscapes) were utilized for gene flow simulations.

The spatial arrangement (and geographic coordinates) of individuals was identical for each population to facilitate comparison among the five resistance surfaces. Four hundred individuals were randomly generated onto the landscape using the "Create Random Points" function in ArcMap. In addition to spatial coordinates for each individual, pair-wise cost matrices were calculated based on raster values using the "cost matrix" function found in the landscape genetics toolbox for ArcGIS (Etherington 2011). This tool sums the resistances of pixels along a Euclidean line connecting two individuals, and then produces a pair-wise cost distance matrix. Euclidean distances between all pairs of individuals were also calculated for Mantel tests using the "Euclidean Distance" function in the toolbox.

Gene Flow Simulations

Individual-based simulations were performed in the program CDPOP v. 1.3 (Landguth & Cushman 2010) using the cost matrices of the simulated landscapes. In total, 10,000 populations were simulated, 500 populations (5 replicates of 100 populations) for each of the 5 resistances (R1, R5, R10, R25, R50) across four habitat area categories (20%, 40%, 60%, 80% suitable habitat). Each simulation started with 400 individuals (200 male, 200 female) characterized at 10 polymorphic microsatellite loci (15 alleles per locus). Dispersal and mating movements occurred according to an inverse square distribution. An inverse square distribution describes a dispersal pattern in which it is far more common for individuals to remain near their birthplace than to disperse long distances. This dispersal distribution typifies many species in nature due to the risks associated with long-distance dispersal (Mayr 1963; Endler 1977). Furthermore, an inverse square dispersal distribution ensures an underlying IBD pattern in all simulated populations. Pair-wise cost distance matrices constrained both mating and dispersal where the maximum movement through suitable habitat was the length of the landscape (6120 m) and decreased according to resistance strength in matrix (R5 = 1224 m, R10 = 612 m, R25 = 244.8 m, R50 = 122.4 m). Mating was sexual with replacement, and each female could have a maximum of five offspring with an average of two. Generations were non-overlapping and results were saved after 500 generations. Statistics were then run on 400 sampled individuals.

Statistical Analysis: Mantel Tests

Mantel tests require pair-wise genetic distances, so we used SPAGEDI v. 1.3 (Hardy & Vekemans 2002) to calculate matrices of inter-individual genetic distances (Rousset's *a*; Rousset 2000), for each simulated population. Matrices of Euclidean, cost, and genetic distances (Table 1.1) were then utilized for full and partial Mantel tests in the R package VEGAN (functions *mantel* and *mantel.partial*; Oksanen *et al*. 2013). Simple Mantel tests compared genetic distance to Euclidean distances, and partial Mantel tests compared genetic and cost distances while controlling for Euclidean distance. This procedure was repeated for each population across resistance value categories (R1, R5, R10, R25, R50) to quantify the ability of partial Mantel tests to separate the effects of cost and Euclidean distances as the relative effect of cost decreased. Statistical significance in all tests was assessed via Pearson's correlation coefficients (*r*) after 9999 permutations.

Landscape Variable Derivation: Ordination

Ordination techniques can incorporate multiple explanatory variables into the analysis, but require individualistic measures of genetic diversity, geography, and costs (Table 1.1). One difficulty in adapting ordination techniques to individual-based landscape genetics is quantification of landscape resistance to gene flow typically relies on inter-individual distances. Multiple techniques exist to quantify cost distances (e.g., least cost path analysis; circuit theory; McRae and Beier 2007), but cost distance calculations often require substantial life history information to properly parameterize a landscape (Spear *et al*. 2010). Information on dispersal itself instead of habitat use is very difficult to obtain, so even with habitat use data, cost distances may not reflect the

factors that affect dispersal (Zeller *et al*. 2012). Therefore, we quantified habitat area around each individual to derive a landscape variable that does not depend on cost distances.

Ordination tests incorporated point estimates of habitat area surrounding each individual derived using following methodology. Circular buffers of 100, 200, and 300 m were initially drawn around each individual point using the "Buffer" function in ArcMap v. 10. Then, the area of matrix habitat was calculated within each buffer and expressed as a proportion (area of matrix/total area in buffer) for use the landscape predictor variable. Point estimates of landscape resistance do not require *a priori* parameterization, which makes them an important complement to analyses that utilize cost distances derived from natural history data or expert opinion.

Statistical Analysis: Ordination

Calculations in dbRDA involve transforming genetic distances (Rousset's *a*) using a principal coordinates analysis (PCoA; Legendre & Anderson 1999) and then applying a normal RDA on the derived PCoA axes. All PCoA axes with positive eigenvalues are retained for analysis. When distances are completely linear, no axes will have negative eigenvalues. Genetic distances, however, are often not linear (Graves *et al*. 2013), so dbRDA effectively removes errors created by non-linear distances to make them appropriate for RDA. As an alternative to PCoA, we also utilized spatial principal components analysis (sPCA; Jombart *et al*. 2008) to derive linear, spatially lagged scores as explanatory variables for a normal RDA.

In population-based analyses, population averages of genetic diversity are often employed as response variables in ordination techniques, but this method is not feasible for individual-based genetics. sPCA incorporates both genetic data and spatial autocorrelation to summarize overall genetic and spatial patterns into independent ordination axes. We chose to use sPCA because this technique removes much of the extraneous variation in genetic data to isolate important patterns and has been successful in isolating landscape effects in empirical individual-based studies (e.g., Robinson *et al*. 2012; Chapter 2). Unlike dbRDA that retains all positive PCoA axes, authors can choose the number of retained ordination axes from sPCA (Jombart *et al*. 2008). Typically, axes with the highest eigenvalues are retained for further analysis, but the exact number often depends on a study's focus and underlying processes affecting gene flow. For simplicity, we retained the first two sPCA axes in all calculations, and derived sPCA lagged scores for genetic explanatory variables. Lagged scores correspond to each individual's position in ordination space and incorporate both genetic diversity and spatial autocorrelation. All sPCA calculations were performed in the R package ADEGENET (Jombart *et al*. 2008).

Full and partial dbRDAs (function *capscale*) and RDAs (function *rda*) were performed in the R package VEGAN (Oksanen *et al*. 2013). Statistical significance for all ordination techniques (dbRDA and RDA) was assessed using the *anova.cca* function in VEGAN; this function is specifically designed to calculate pseudo *F*-ratios, variance components, and *p*-values from ordination methods (Oksanen *et al*. 2013). Since the total variance explained by ecological and geographic variables is typically very small (1.20 – 12.35% in our simulated datasets), we extracted pseudo *F*-ratios because like Pearson

correlation coefficients, higher *F*-ratios indicate a stronger relationship between landscape or geographic variables and genetic variation.

Statistical Performance

In this study, statistical performance was first defined by counting the number of times each test correctly identified the underlying pattern of IBD or IBR (hereby termed accuracy; Table 1.1). Because all landscapes were characterized by an underlying IBD pattern, a correct test occurred when a full test was significant in the R1-R50 landscapes. In the IBR case (R5-R50 landscapes), accuracy of a partial test (in which geographic distances were controlled) was defined by how many significant partial tests occurred out of 100 simulated populations. Along with accuracy, we also counted how often each test did not correctly identify the appropriate mechanism driving gene flow (type I or II errors; Table 1.1). Type I errors occur only in the R1 (IBD only) scenario when partial tests are significant for IBR despite no underlying IBR. In contrast, type II errors were defined as the inability to detect the effect of matrix habitat in the IBR landscapes (R5-R50).

We also assessed the effect of landscape heterogeneity (percent suitable habitat, landscape configuration metrics, and resistance strength) on the strength of detected IBR relationship (i.e., explanatory power of the model). General linear models (glm) for each partial test incorporated either F-ratios (ordination techniques) or Pearson correlation coefficients (Mantel tests) calculated from the IBR scenarios (R5-R50) as explanatory variables. Predictor variables included six variables: percent suitable habitat, clumpy, and resistance as well as the residuals of regressions between percent suitable habitat and

correlation length or patch cohesion. Separating configuration (i.e., correlation length and patch cohesion) from habitat amount (percent suitable habitat) effects can be difficult because these landscape variables are highly correlated (all Mantel $r > 0.866$; $p < 0.001$), but Cushman *et al*. (2012) showed that configuration metrics explained additional variance in Pearson correlation coefficients. Therefore, percent suitable habitat was regressed against patch cohesion and correlation length to derive residuals that represent the variance in patch cohesion and correlation length not explained by percent suitable habitat. As our total dataset contained 8,000 populations (excluding IBD cases), we randomly selected 20 populations out of the total 100 within each landscape/resistance combination to minimize errors created by large sample size (final n = 1,600).

Model selection used an information criterion approach (Burnham & Anderson 2002) to identify the best model among 31 candidate models. Model ranking occurred according to Burnham & Anderson (2002) where models with the lowest Akaike Information Criterion (AIC) and a ΔAIC value < 2.0 were the best models. We calculated AIC weights to examine how likely the top model is the best model among all candidate models. Final parameter estimates for resistance and landscape variables were calculated through model averaging of all top models, and parameters that do not include zero explained significant variation within Mantel *r* or *F*-ratios. All glm and AIC procedures were performed in R.

Results

Both partial Mantel tests and dbRDA identified IBD in all populations (100%) whereas RDA could not explicitly test for IBD because sPCA axes explicitly incorporate

both spatial autocorrelation and genetic variation, making it unable to isolate IBD in a simple test. Unlike IBD detection, where both partial Mantel tests and dbRDA consistently yielded high accuracy rates, IBR detection was highly variable across landscapes and resistance strengths. In the 20% suitable habitat landscapes, partial Mantel tests performed better (53-100% accuracy) than ordination methods (4-94% accuracy), but then were less accurate than ordination as the amount of suitable habitat increased. Type I error rates (i.e., detecting IBR when the resistance of matrix habitat was 1), were lower (5-26%) than type II error rates (17% - 96%) in Mantel tests. Ordination approaches also had higher type I errors, particularly in dbRDA (25-99%; details below; Figure 1.2). Type II errors (i.e., failure to detect IBR when resistance of matrix habitat was greater than 1) occurred in all landscapes and resistances with relatively high frequencies for all tests. High type II error rates were not unexpected given that intrapopulation levels of genetic structure are often quite small. Resistance strength, percent suitable habitat, and landscape configuration influenced all statistical tests' abilities to detect the effect of matrix habitat.

The effects of landscape configuration and resistance strength on IBR detection accuracy were different for each test. Maximum performance in partial Mantel tests occurred within the 20% suitable habitat at resistance 50 where significant IBR was detected in 82-99% of populations. Correlation coefficients describing relationships between cost distances and genetic differentiation were low overall, ranging from -0.0487-0.125. The two top ranked glm models (summed AIC weight = 0.9991; Table 1.2) included percent suitable habitat, resistance strength, correlation length, and clumpy (all $p < 0.005$). Pearson correlation coefficients were negatively correlated with percent

suitable habitat, correlation length, and clumpy and positively correlated with resistance strength (Table 1.3). Taken together, these results show that partial Mantel tests detected IBR more often in landscapes with low percent suitable habitat, very high resistance of matrix habitat, and high fragmentation of suitable habitat (low clumpy and correlation length).

Overall, dbRDA had very high type I error rates where IBR was falsely detected in 62-99% of R1 populations (Figure 1.2). In IBR landscapes (R5-R50), dbRDA rarely detected IBR as frequently as the R1 landscapes (Figure 1.3). Accuracy within IBR landscapes was highly variable among resistances and amount of suitable habitat, but highest performance generally occurred within the 20% and 40% suitable habitat landscapes. The three top glm models (summed AIC weight = 0.9304) indicated that F-ratios in R5-50 populations were positively associated with clumpy and negatively related to percent suitable habitat and correlation length (Table 1.3). All three variables remained significant after model averaging. Resistance and patch cohesion were not found to be significant factors in explaining F-ratios (Table 1.2).

Like dbRDA, RDA with sPCA spatially lagged scores had higher type I errors than partial Mantel tests (Figure 1.2). Type I errors in RDA ranged from 25-60%, which was similar to accuracy rates in IBR scenarios (20-68%; Figure 1.3). However, IBR landscapes always had higher numbers of significant tests than the IBD case. The three top glm models included percent suitable habitat, clumpy, and correlation length (summed AIC weight = 0.6259; Table 1.2). Only clumpy and the intercept remained significant after model averaging (Table 1.3), which is likely due to over half of competing models having ΔAICs from 3.0 to 9.0. Clumpy was positively associated with

F-ratios whereas percent suitable habitat and correlation length were negatively correlated with F-ratios in RDA. Taken together, the results for dbRDA and RDA are concordant, and show that IBR detection is highest in fragmented landscapes with a few clumped patches (low correlation length and high clumpy) and low-intermediate amounts of suitable habitat.

Discussion

Ordination techniques possess numerous characteristics that were predicted to enhance differentiation between IBD and IBR models of gene flow in individual-based studies over more common distance-based methods like Mantel tests. Ordination methods are able to overcome the linearity assumptions that characterize distance-based methods, offer flexibility in the type and number of explanatory variables, and provide a suite of statistical outputs that permit thorough interpretation of results. Despite predictions that ordination methods would outperform Mantel tests, our data show that no one test was uniformly superior to any other. Across all landscapes, both partial Mantel tests and dbRDA were effective at detecting IBD regardless of the strength of IBR. The need to transform genetic data for RDA prevented complete isolation of IBD and IBR, making it impossible to explicitly test for IBD. In landscapes that were characterized by both IBD and IBR, IBR detection was highly variable across tests. Mantel tests use pairwise distances that have considerable noise, likely resulting in decreased power to detect IBR. In contrast, ordination techniques are more powerful tests (i.e., can detect IBR more often), but suffer from false detections of IBR likely created by variable

transformations. Landscape composition also had a powerful impact on all test statistics. Maximum performance occurred in landscapes where IBR was strongest (highly fragmented with low suitable habitat). Ordination methods were more robust to changes in the underlying landscape than Mantel tests, suggesting that ordination methods might provide more meaningful conclusions about gene flow when little is known about the resistance of the landscape.

Performance of Mantel Tests and Ordination Techniques

We found that both Mantel tests and dbRDA were highly effective at detecting IBD because they always detected IBD when it was present regardless of the strength of IBR. These high accuracy rates indicate that investigators can be confident that significant Mantel tests or dbRDA are strong evidence for IBD in a natural population. RDA, in contrast, could not test for IBD because sPCA does not explicitly control for IBD (Jombart *et al*. 2008), resulting in spatially lagged scores that reflect both IBR and IBD. IBD and IBR typically occur simultaneously in nature, so verification of IBD is a critical step before isolating the effects of IBR. Unlike the high accuracy and power in IBD tests, IBR detection was highly variable across tests and landscape configurations.

Mantel tests, the most common test in individual-based landscape genetics, generally had low accuracy rates for IBR tests, failing to detect IBR when it was present. These high type II error rates suggest that Mantel tests' dependence on pair-wise genetic, cost, and Euclidean distances may reduce power to detect IBR. Indeed, Legendre and Fortín (2010) found that Mantel tests often suffer from reduced power as compared to ordination techniques due to data transformations into pair-wise distances (Legendre &

Fortín 2010). If Mantel tests inherently have reduced power to detect IBR, then the subtle genetic differentiation present in individual-based studies may be too low to consistently detect IBR. The inability of Mantel tests to detect intrapopulation levels of genetic differentiation is concerning because Mantel tests may not detect IBR in empirical studies with levels of genetic differentiation similar to this study. For example, failure to identify genetic differentiation caused by anthropogenic land use could be a critical problem when trying to inform management agencies about how fragmentation impacts populations of conservation concern. Mantel tests can detect barriers if there is adequate lag time (Landguth *et al*. 2010), but recent anthropogenic barriers may not result in sufficient genetic differentiation to detect IBR. In cases where Mantel tests do not detect a barrier, we recommend use of more powerful tests and field data to better understand anthropogenic barriers to gene flow. Alternatively, if Mantel tests do detect a barrier, researchers can be confident that IBR is present because Mantel tests rarely detected IBR in IBD only landscapes within our simulations.

Despite their low accuracy rates in IBR detection, Mantel tests rarely detect IBR when only IBD is present (i.e., type I errors). This finding illustrates the utility of Mantel tests for individual-based landscape genetics. The low type I error rates found in this study are in contrast to findings for population-based simulation studies, which have found high rates of false IBR detection using Mantel tests (e.g., Balkenhol *et al*. 2009; Guillot & Rousset 2013). High type I errors for Mantel tests in population-based studies are likely due to greater genetic differentiation between population pairs and substantially fewer pair-wise comparisons than individual-based studies, both of which increase power to detect IBR (and the potential for type I errors). Individual-based simulation studies

have also recorded greater type I errors than this study, especially in complex landscapes with 3 or more habitat types (Graves *et al.* 2013). In such complex landscapes, a type I error can stem from false significance of an IBR test in an IBD landscape or from detection of an incorrect landscape factor driving gene flow. More ways to generate type I errors in complex landscapes leads to higher overall type I error rates (Graves *et al.* 2013).

Our results suggest that ordination techniques (dbRDA and RDA) detect IBR more often than partial Mantel tests based on the higher number of significant partial tests in most IBR landscapes. The lower type II error rates in ordination techniques suggest they have higher power to detect IBR patterns than partial Mantel tests, which can be highly useful in situations where genetic differentiation is low, as is common in individual-based studies. For example, anthropogenic changes to once continuous landscapes (e.g., roads or land conversion) require multiple generations of reduced gene flow to produce genetic differentiation within a population (Landguth *et al*. 2010). Ordination techniques could potentially detect such contemporary barriers faster than Mantel tests because they can detect weaker IBR. However, one critical caveat of ordination techniques is their propensity to detect IBR even in populations that only exhibit IBD (i.e., type I errors). Both dbRDA and RDA exhibited high type I errors, which indicates they have difficulty removing the effects of IBD from IBR.

Type I errors were highest in dbRDA, a likely result of the transformation of genetic distances. Pair-wise genetic distances are transformed using a PCoA, and then all resultant PCoA axes with positive eigenvectors are retained for analysis. In essence, PCoA removes the often high variability in raw genetic distances between individuals, so

the reduced noise within genetic data may inflate type I error rates. PCoA is based on linear distances, so as genetic distances become less linear due to IBR, negative PCoA axes will occur. Thus, the amount of variation explained by PCoA axes is maximized within the R1 landscape because genetic distances are only correlated with Euclidean distance, making the amount of variation contained within PCoA axes very high. When IBR occurs, genetic distances are no longer solely correlated with Euclidean distance, so the PCoA attempts to force the non-Euclidean genetic distances into linear axes. Any remaining variation goes into negative PCoA axes and is not used in the dbRDA, so the amount of variation explained in genetic distances is less than in the R1 landscape. Though PCoA transformations are designed to linearize and reduce noise in genetic data, dbRDA may not produce valid results in individual-based studies because it cannot effectively distinguish between IBD and IBR due to high type I errors.

Like dbRDA, RDA with sPCA lagged scores exhibited high type I errors, which also is likely related to transformation of genetic data into ordination axes. Further investigation is warranted to understand if including more or fewer ordination axes impacts performance in sPCA (and other similar techniques that would collapse genetic variation into ordination axes). In this study, we retained the first two sPCA axes because they explained the most variance and their eigenvalues were usually well differentiated from other axes. Even though sPCA summarizes patterns in spatial autocorrelation and genetic diversity, it does not explicitly control for it, so IBD can contribute to multiple sPCA axes (Jombart *et al*. 2008). As a result, ordination axes may just reflect IBD instead of patterns in IBR, so use of sPCA and RDA requires individual inspection of each axis to identify patterns that may be associated with IBD and IBR.

Screeplots from sPCA aid in this process because they provide a visual representation of the amount of spatial autocorrelation within each axis (Jombart *et al*. 2008). Essentially, there is no rule for the number of sPCA axes to retain for genetic analysis, and each axis should be inspected individually to help understand how IBD and IBR impact genetic variation.

Although ordination techniques are more powerful tests (i.e., can detect IBR more often), removing IBD from IBR is difficult when using ordination-based genetic variables. Population-based studies often do not suffer from this problem because they can utilize population-specific variables such as allele frequencies and connectivity indices (see Balkenhol *et al*. 2009). Population-specific allele frequencies and connectivity indices cannot be calculated for individuals leaving few alternatives to partial Mantel tests in individual-based studies. To date, focus in landscape genetic methods has generally been placed on comparing or developing methods that are alternatives to partial Mantel tests in population-based studies (e.g., Robinson *et al*. 2012; Wang 2013). Many of these population-based techniques offer considerable promise in individual-based landscape genetics, but development of proper genetic variables for individual-based techniques remains problematic. Therefore, we recommend exploration into alternative individual-based variables that could better disentangle IBR from IBD to maximize the utility of ordination and regression statistics in landscape genetics.

Influence of Landscape Composition on Partial Statistics

While intrinsic factors (i.e., variable type) of each test certainly contributed to accuracy and power, the underlying landscape also had a powerful impact on each tests'

ability to detect IBR. Landscapes with high fragmentation and low suitable habitat had the highest performance indices (highest accuracy and test statistics and lowest error rates) across all tests. The high performance in the most extreme cases of fragmentation and habitat amount supported our prediction that highly fragmented landscapes with low suitable habitat would produce the strongest amount of genetic differentiation, making IBR easier to detect. Despite that maximum performance occurred in the same fragmented landscapes with low suitable habitat, partial Mantel and ordination test statistics were sensitive to different landscape metrics (i.e., percent suitable habitat, correlation length, patch cohesion, clumpy, and resistance strength).

Mantel correlation coefficients were significantly associated with all landscape metrics except patch cohesion, indicating Mantel tests are highly sensitive to changes in the underlying landscape. Mantel tests detected IBR more often than ordination techniques only in the landscapes with the most extreme levels of fragmentation (i.e., low correlation length), suitable habitat, and matrix resistance. Correlation length was particularly important for explaining Mantel correlation coefficients because correlation length is an indicator of the extensiveness of suitable habitat. Mantel tests rely on pair-wise distance calculations, so measures of landscape resistance and genetic differentiation are highly dependent on how much suitable habitat occurs between two individuals. As correlation length decreases, the amount of suitable habitat between two individuals also decreases, which in turn elevates genetic distances. Therefore, a strong Mantel correlation coefficient will occur in landscapes with small correlation lengths (see Cushman *et al*. 2012). Another important factor in explaining Mantel correlation coefficients was resistance strength of matrix habitat where a positive relationship was

recorded. Higher resistance strength increases genetic differentiation, so this relationship was not surprising, but it does raise an important issue for use of Mantel tests in empirical studies. Estimating resistances can be difficult without proper life history (Spear *et al*. 2010; Zeller *et al*. 2012), so the degree of error in Mantel tests will be unknown. Based on the fairly limited situations where Mantel tests have high accuracy rates, partial Mantel tests should be paired with other analyses like ordination techniques to help distinguish between a type II error and IBD.

Maximum performance (i.e., highest accuracy rates and F-ratios) of ordination methods occurred within the same fragmented, low suitable habitat landscapes as Mantel tests, but clumpy was the most important landscape variable. Clumpy had the biggest impact on strength of IBR in both dbRDA and RDA, which likely reflects the buffer calculation utilized as landscape variables. In more clumped landscapes, percent suitable habitat within a buffer was similar for many individuals. One critical advantage in ordination techniques was that resistance strength of the matrix had no impact on test statistics in either dbRDA or RDA. Ordination techniques do not require accurate parameterization of landscape resistance and are robust to changes in landscape resistance, so ordination techniques could be particularly useful for species lacking relevant life history data (e.g., presence-absence, mark-recapture, path analysis; Zeller *et al*. 2012) to parameterize resistance landscapes. Ordination techniques offer considerable promise for empirical studies, and we encourage further study with more complex landscapes and natural populations to understand performance of ordination techniques in individual-based landscape genetics.

Conclusions

Investigation of IBR using ordination techniques will probably be most informative when paired with simulated populations that exhibit IBD to quantify type I errors. In particular, simulations of populations under IBD (i.e., null hypothesis) with geographic locations, landscape configuration, and genetic diversity that mimic the study population can aid in characterizing how often statistics produce erroneous conclusions about gene flow (Chapter 2). Simulations should closely mirror the empirical dataset because deviating from observed landscape configuration, genetic data parameters, or sampling schemes can have drastic impacts on statistical results (e.g., Cushman *et al*. 2012; Cushman *et al*. 2013). Interpretation of simulated populations in an individual-based framework could follow several paths, but we recommend quantifying type I errors for each proposed landscape hypothesis and only accepting factors that were not detected in IBD landscapes at high frequencies (e.g., 5%). Simulating populations under IBR and comparing them to empirical datasets is another possibility, but meaningful simulations require considerable life history data (i.e., population-specific demographics) that is largely unavailable for many species.

Despite the risks of type I errors with ordination tests, they are potentially more versatile than Mantel tests based on their ability to reduce noise within datasets and utility in a wide variety of landscape genetic scenarios. Partial Mantel tests are likely to perform well in species with strong habitat specialization or patchy distribution (e.g., Schwartz *et al*. 2009; Wasserman *et al*. 2010; Shafer *et al*. 2012). Our results indicate that partial Mantel tests may not be accurate for species that utilize a variety of habitats or are continuously distributed. For these species, estimating resistance can be difficult even with the inclusion of individual movements that relate to single landscape factors

(e.g., Zeller *et al*. 2012). In cases where the resistances of landscape factors are unknown and population boundaries are difficult to define, solely applying Mantel tests is unlikely to be an effective method to understand genetic connectivity. Combining Mantel tests and ordination techniques to evaluate specific appropriate hypotheses, synthesizing the results across approaches to facilitate comprehensive interpretation, and using simulations to evaluate significance of test statistics may be the best way to glean meaningful conclusions about gene flow in individual-based landscape genetics.

Figure 1.1. Multifractal landscapes generated for this study where 400 individuals were randomly placed on each landscape. Each landscape consisted of two landscape categories: suitable (light grey) and matrix (black) habitat. The amount of suitable habitat (P) varied from 20% to 80% with five replicates within each P category. Pixels of suitable habitat were given a resistance value of 1 whereas matrix habitat was given one of four values (5, 10, 25, and 50) across simulations for a total of 80 simulated landscapes.

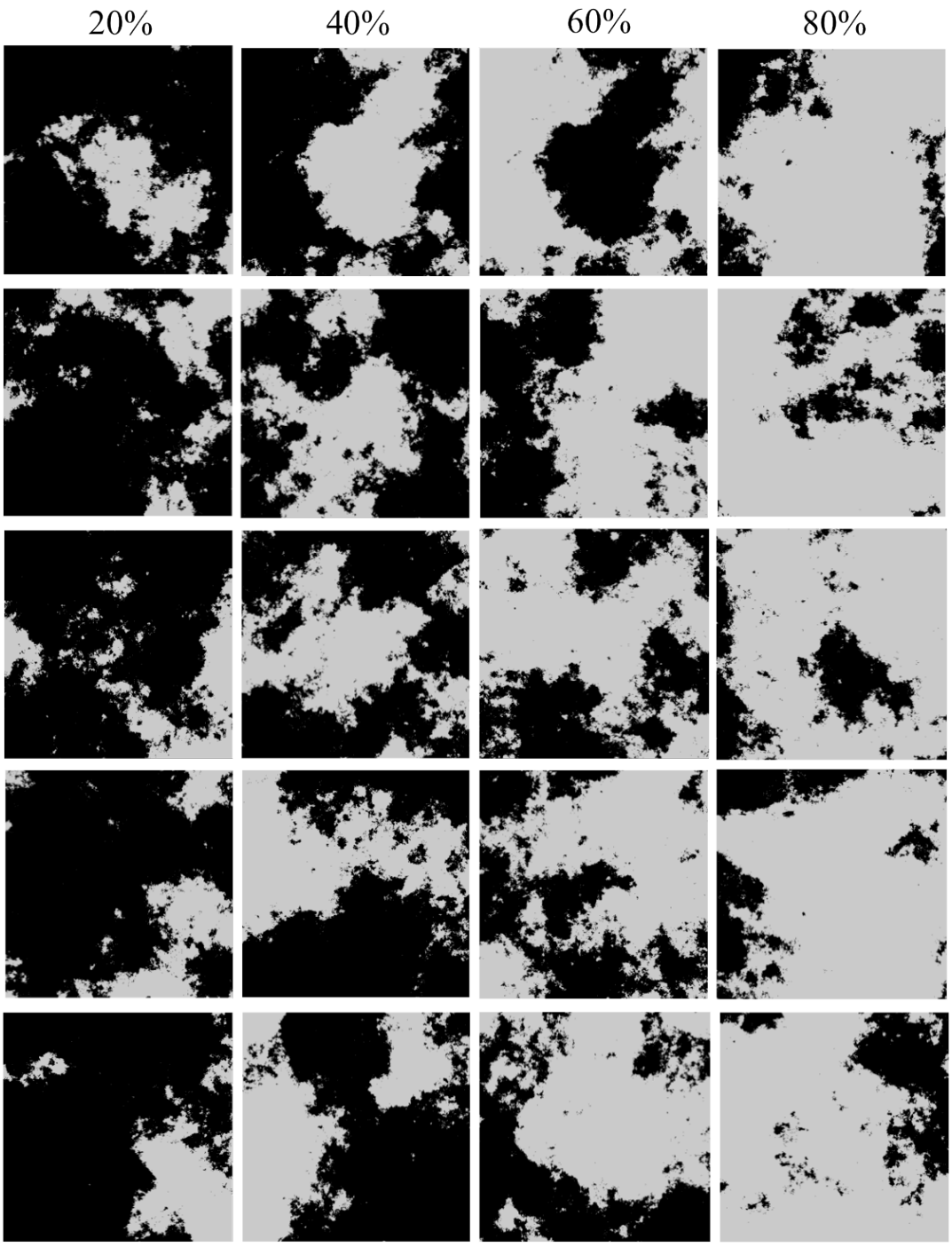


Figure 1.2. Frequency of type I errors in partial Mantel tests, dbRDA, and RDA in simulated landscapes. IBR was incorrectly detected most often in dbRDA tests, and least often in partial Mantel tests. Error bars correspond to standard errors calculated across the 1,600 simulated populations.

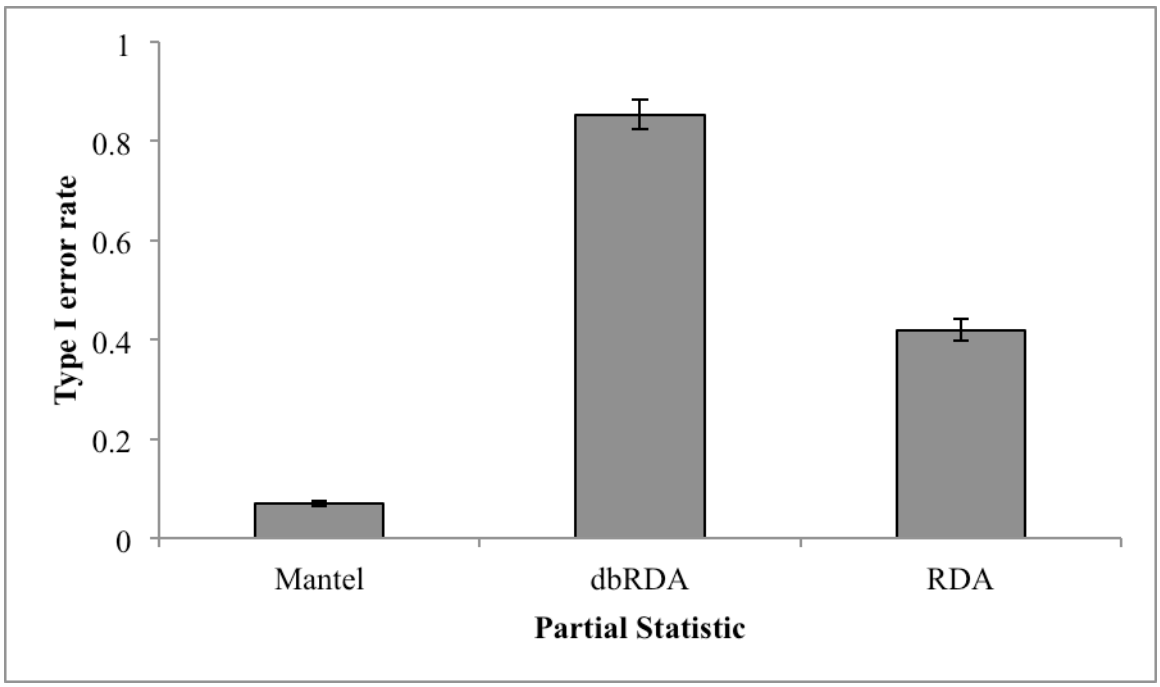
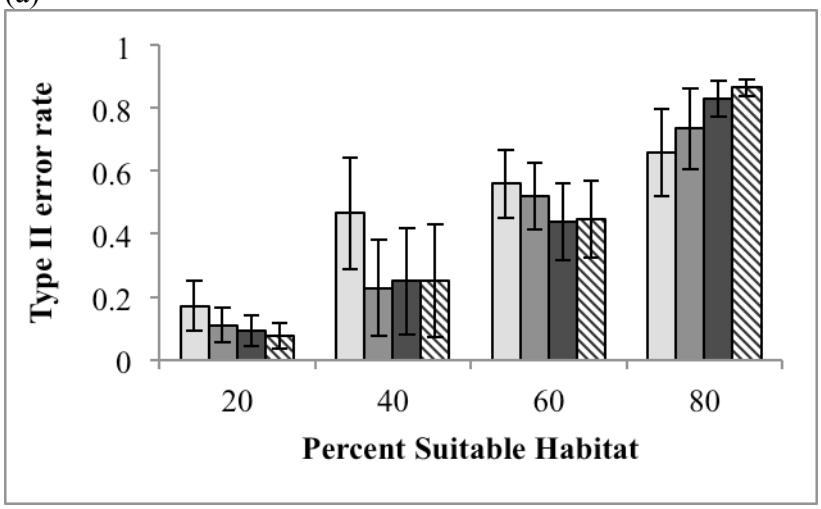
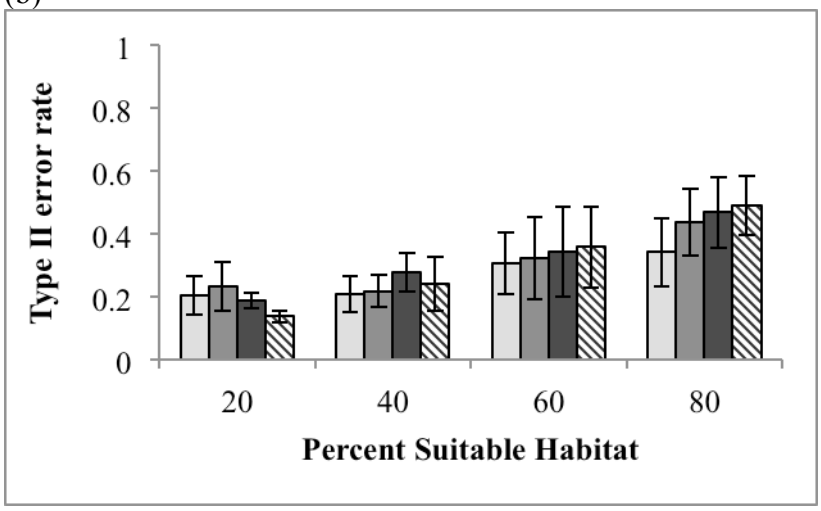


Figure 1.3. Frequency of type II errors in partial Mantel tests (a), dbRDA (b), and RDA (c) in simulated landscapes. Five replicate landscapes were simulated for each percent suitable habitat (20%, 40%, 60%, 80%), at four matrix habitat resistance levels (R5 – light gray bars; R10 – gray bars; R25 – dark gray bars; R50 – striped bars). One hundred populations were simulated for each of the 80 total landscapes. Type II errors were calculated as the number of populations (out of 100) in which IBR was not detected in our simulated landscapes, and corresponding error bars indicate standard errors calculated across five replicate landscapes. In general, type II errors increased with amount of suitable habitat but were similar across resistance levels for all tests.

(a)



(b)



(c)

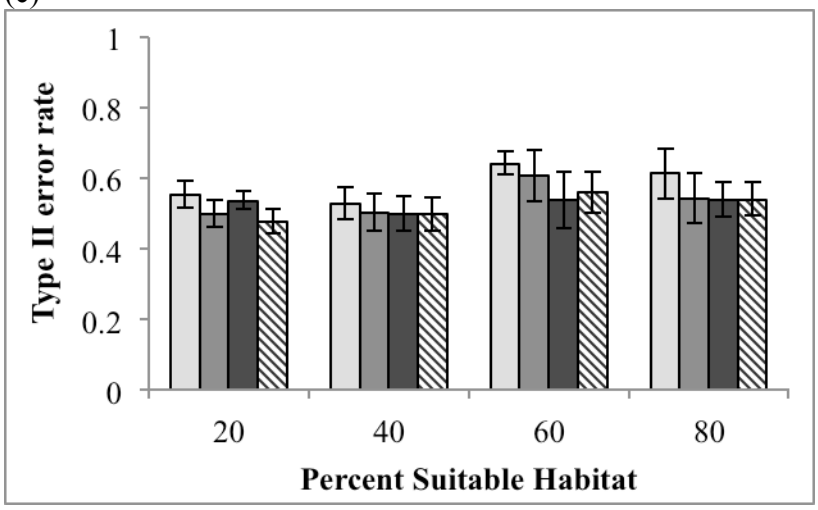

Type II error rate
1
0.8
0.6
0.4
0.2
0
20
40
60
80
Percent Suitable Habitat

Table 1.1. Summary of variables calculated for partial Mantel tests, dbRDA, and RDA analyses. Partial Mantel tests use pair-wise distances for all variables. dbRDA uses genetic distance as the response variable, and individualistic measures of ecological and geographic data as predictors. Genetic data for RDA was derived from a spatial principal components analysis (sPCA). All three tests had three indices of performance (accuracy, type I, and type II error rates) calculated based on 100 simulated populations per landscape.

	Partial Mantel	dbRDA	RDA
Genetic data	Pair-wise genetic distance	Pair-wise genetic distance	sPCA spatially lagged scores
Landscape data	Pair-wise cost distance	% Suitable Habitat in Buffer	% Suitable Habitat in Buffer
Geographic data	Pair-wise Euclidean distance	Geographic coordinates	Geographic coordinates
Accuracy	Proportion of populations that had significant IBD or IBR tests when IBD or IBR was present		
Type I error rate	Proportion of populations where tests were significant for IBR when only IBD was present		
Type II error rate	Proportion of populations where tests were not significant for IBR when IBR was present		

Table 1.2. Top six models that explained variation in partial Mantel *r*s (a), dbRDA pseudo F-ratios (b), and RDA pseudo F-ratios. Models are ranked according the lowest AIC value, and all those with a ΔAIC < 2.0 are considered top models (bold).

(a)

Model	AIC	ΔAIC	Weight
%SH + Clumpy +Resistance + ResCL	**-7638.7**	**0.0**	**0.4996**
%SH + Clumpy + Resistance + ResCL + ResCoh	**-7638.7**	**0.0**	**0.4996**
%SH + Resistance + ResCL	-7625.4	13.3	0.0006
%SH + Resistance + ResCL + ResCoh	-7623.4	15.3	0.0002
%SH + Clumpy + ResCL	-7603.2	35.5	0.0000
%SH + Clumpy + ResCL + ResCoh	-7603.1	35.6	0.0000

(b)

Model	AIC	ΔAIC	Weight
%SH + Clumpy + ResCL	**2783.6**	**0.0**	**0.4883**
%SH + Clumpy + Resistance + ResCL	**2785.0**	**1.4**	**0.2424**
%SH + Clumpy + ResCL + ResCoh	**2785.6**	**2.0**	**0.1796**
%SH + Clumpy + Resistance + ResCL + ResCoh	2787.0	3.4	0.0892
Clumpy + ResCL	2799.5	15.9	0.0002
Clumpy + Resistance + ResCL	2800.8	17.2	0.0000

(c)

Model	AIC	ΔAIC	Weight
%SH + Clumpy + ResCL	**11437**	**0.0**	**0.3170**
%SH + Clumpy + Resistance + ResCL	**11438**	**1.0**	**0.1923**
%SH + Clumpy +ResCL + ResCoh	**11439**	**2.0**	**0.1167**
%SH + Clumpy + Resistance + ResCL + ResCoh	11440	3.0	0.0707
%SH + Clumpy	11440	3.0	0.0707
%SH + Clumpy + Resistance	11441	4.0	0.0429

%SH: percent suitable habitat; ResCL: residuals of linear regression between correlation length and percent suitable habitat; ResCoh: residuals of linear regression between patch cohesion and percent suitable habitat

Table 1.3. Model averaged parameters for the top glm models in partial Mantel tests, dbRDA, and RDA. Only parameters that were shared between models are reported (i.e., models with an ΔAIC < 2.0). When 95% confidence intervals did not include zero, parameters were considered significant (bold).

	Estimate	SE	Upper CI	Lower CI
Partial Mantel tests				
%SH	**-0.0005**	**5.9953E-05**	**-0.0005**	**-0.0006**
Clumpy	-0.1379	0.0387	-0.0621	-0.2137
Resistance	**0.0002**	**3.1667E-05**	**0.0003**	**0.0001**
ResCL	**0.0006**	**0.0005**	**0.0008**	**0.0003**
Intercept	**0.1693**	**0.0341**	**0.2384**	**0.1021**
dbRDA				
%SH	**-0.9376**	**0.0013**	**-0.9351**	**-0.9401**
Clumpy	**7.1272**	**1.5802**	**10.2244**	**4.0299**
ResCL	**0.0096**	**0.0040**	**0.0174**	**0.0019**
Intercept	**-4.7656**	**1.1481**	**-2.5153**	**-7.0159**
RDA				
%SH	-0.0195	0.0132	0.0063	-0.0453
Clumpy	**59.3786**	**16.3289**	**91.3832**	**27.3739**
ResCL	0.0424	0.0412	0.1232	-0.0384
Intercept	**-49.0327**	**14.9210**	**-19.7876**	**-78.2779**

%SH: percent suitable habitat; ResCL: residuals of linear regression between correlation length and percent suitable habitat; SE: unconditional standard errors; Upper and Lower CIs: 95% confidence interval boundaries

Literature Cited

Balkenhol N, Waits LP, Dezanni RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, **32**, 818–830.

Bruggeman DJ, Wiegand T, Fernandez N (2010) The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). *Molecular Ecology*, **19**, 3679–3691.

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multi-locus molecular markers. *Molecular Ecology*, **15**, 2833-2843.

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology*, **19**, 3592-3602.

Cushman SA, Lewis JS (2010) Movement behavior explains genetic differentiation in black bears. *Landscape Ecology*, **25**, 1613-1625.

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

Cushman SA, Shirk A, Landguth EL (2012) Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex habitats. *Landscape Ecology*, **27**, 369-380.

Cushman SA, Shirk A, Landguth EL (2013) Landscape genetics and limiting factors. Conservation Genetics, 14, 263-274.

Endler JA (1977) Geographic variation, speciation and clines. Princeton University Press, Princeton, New Jersey, USA.

Etherington TR (2011) Python based GIS tools for landscape genetics: visualizing relatedness and measuring landscape connectivity. *Methods in Ecology and Evolution*, **2**, 52-55.

Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493-505.

Gardner RH (1999) RULE: a program for the generation of random maps and the analysis of spatial patterns. In: Klopatek JM, Gardner RH (eds) Landscape ecological

analysis: issues and applications. Springer, New York.

Graves TA, Beier P, Royle JA (2013) Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Molecular Ecology*, **22**, 3888-3903.

Guillot G, Mortier F, Estoup A (2005) Geneland: A program for landscape genetics. *Molecular Ecology Notes*, **5**, 712-715.

Guillot G, Rousset F (2013) Dismantling Mantel tests. *Methods in Ecology and Evolution*, **4**, 336-344.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405.

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92-103.

Keyghobadi N (2007) The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology*, **85**, 1049–1064.

Kierepka EM (2014) Landscape genetics of the American badger (*Taxidea taxus*): understanding challenges with elusive species. *Ph.D. thesis*. University of Wisconsin-Milwaukee, Milwaukee, WI.

Landguth EL, Cushman SA (2010) CDPOP: an individual-based, cost-distance spatial population genetics model. *Molecular Ecology Resources*, **10**, 156–161.

Landguth EL, Cushman SA, Luikart G, Murphy MA, Schwartz MK, McKelvey KS (2010) Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, **19**, 4179-4191.

Landguth EL, Fedy BC, Oyler-McCance SJ, Garey AL, Emel SL, Mumma M, Wagner HH, Fortín M, Cushman SA (2012) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Molecular Ecology Resources*, **12**, 276-284.

Latch EK, Dharmarajan G, Glaubitz J, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295-302.

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multiple species responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1-24.

Legendre P, Fortín M (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831-844.

Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge, Massachusetts, USA.

McGarigal K, Cushman SA (2002) Comparative evaluation of experimental approaches to the study of habitat fragmentation effects. *Ecological Applications*, **12**, 335–345.

McGarigal K, Cushman SA, Neel MC et al. (2002) FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: http://www.umass.edu/landeco/fragstats.html.

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

McRae BH (2006) Isolation-by-resistance. *Evolution*, **60**, 1551-1561.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839-2846.

Neel MC, Cushman SA, McGarigal K (2004) Behavior and stability of landscape metrics across controlled gradients of landscape structure. Landscape Ecology, 19, 435–455.

O'Brien D, Manseau M, Fall A, Fortin MJ (2006) Testing the importance of spatial configuration of winter habitat for woodland caribou: an application of graph theory. *Biological Conservation*, **130**, 70-83.

Oksanen J, Guillaume-Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Peter Solymos, Stevens MHH, Wagner H (2013) VEGAN: Community Ecology Package. R package version 2.0-5. http://CRAN.R-project.org/package=vegan.

Oyler-McCance SJ, Fedy BC, Landguth EL (2013) Sample design effects in landscape genetics. *Conservation Genetics*, **14**, 275-285.

Pritchard JK, Stephens P, Donnelly P (2000) Inference of population genetic structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Reding DM, Cushman SA, Gosselink TE, Clark WR (2013) Linking movement behavior and fine-scale genetic structure to model landscape connectivity for bobcats (Lynx rufus). *Landscape Ecology*, **28**, 471-486.

Revilla E, Wiegand T (2008) Individual movement behavior, matrix heterogeneity, and the dynamics of spatially structured populations. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19120-19125.

Robinson SJ, Samuel MD, Lopez DL, Shelton P (2012) The walk is never random: subtle landscape effects shape gene flow in a continuous white-tailed deer population in the Midwestern United States. *Molecular Ecology*, **21**, 4190-4205.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58-62.

Schumaker NH (1996) Using landscape indices to predict habitat connectivity. *Ecology* 77:1210–1225.

Schwartz MK, Copeland JP, Anderson NJ, Squires JR, Inman RM, McKelvey KS, Pilgrim KL, Waits LP, Cushman SA (2009). Wolverine gene flow across a narrow climatic niche. *Ecology*, **90**, 3222–3232.

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441-452.

Shafer AB, Northrup JM, White KS, Boyce MS, Coté SD, Coltman DW (2012) Habitat selection predicts genetic relatedness in an alpine ungulate. *Ecology*, **93**, 1317-1329.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

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

Walpole AA, Bowman J, Murray DL, Wilson PJ (2012) Functional connectivity of lynx at their southern range periphery in Ontario, Canada. *Landscape Ecology*, **27**, 761-773.

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, **67**, 3403-3411.

Wasserman TN, Cushman SA, Schwartz MK, Wallin DO (2010) Spatial scaling and multi-model inference in landscape genetics: *Martes americana* in northern Idaho. *Landscape Ecology*, **25**, 1601-1612.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Zeller KA, McGarigal K, Whiteley AR (2012) Estimating landscape resistance to gene flow: a review. *Landscape Ecology*, **27**, 977-797.

CHAPTER 2: FINE-SCALE LANDSCAPE GENETICS OF THE AMERICAN BADGER (*TAXIDEA TAXUS*): DISENTANGLING LANDSCAPE EFFECTS AND SAMPLING ARTIFACTS IN AN ELUSIVE SPECIES

Introduction

Landscape genetics has become an increasingly important tool for conservation and management by identifying landscape factors that influence genetic connectivity across heterogeneous landscapes. Maintaining genetic connectivity across landscapes is critical for overall population viability (Hanski 1999; Johansson *et al*. 2007), so landscape genetics can provide information to help understand and potentially mitigate the effects of land use change in natural populations (e.g., fragmentation, habitat loss, anthropogenic disturbance; Segelbacher *et al*. 2010). To understand how land use change may impact genetic connectivity within a focal species, researchers must first identify landscape features that are potentially correlated with gene flow. These landscape genetic hypotheses are typically based on variables that predict habitat preferences or occupancy in field-based investigations (e.g., telemetry or presence-absence studies). This method of hypothesis development has been successful in landscape genetics, and multiple studies have shown that important landscape features in field studies such as land cover (Goldberg & Waits 2010; Garroway *et al*. 2011), climatic conditions (Row *et al*. 2014), and anthropogenic barriers (Blanchong *et al*. 2008; Latch *et al*. 2011) also translate to strong correlations with genetic differentiation. When field data corroborates correlations from landscape genetics (e.g., avoided habitats also prevent gene flow; Shafer *et al*. 2012), meaningful conclusions about genetic connectivity can be drawn and used to develop sound conservation and management plans.

Although landscape genetics has certainly proven to be a robust technique for species with extensive field data (e.g., Cushman *et al*. 2006; Schwartz *et al*. 2009; Shafer *et al*. 2012), application to elusive species can be difficult. Two main challenges arise when working with elusive species in landscape genetics: 1) parameterization of landscape genetic hypotheses without relevant life history data, and 2) deciphering between spurious statistical results and true correlations between patterns of gene flow and landscape variables. Ideally, investigators should develop specific hypotheses about how landscape heterogeneity impacts gene flow in a focal species and parameterize models based on *a priori* hypotheses (e.g., Balkenhol *et al*. 2009; Anderson *et al*. 2010). For many elusive species, however, we lack information about processes that might affect gene flow such as habitat preferences, distribution, and population structure within a specific study area. Most commonly, expert opinion has been used as a proxy for relevant field data (Zeller *et al*. 2012). Using expert opinion to parameterize landscape genetic models has been criticized because it is inherently subjective and difficult to assess accuracy without corresponding field data (Spear *et al*. 2010; Zeller *et al*. 2012). Without separate field-based data to corroborate landscape genetic results, models based on expert opinion can make distinguishing between an erroneous result and an actual landscape effect difficult for investigators. Spurious correlations in popular statistics like Mantel tests are relatively common even with accurate parameterization (Cushman & Landguth 2010; Graves *et al*. 2013), so utilizing expert opinion in elusive species requires additional steps to help alleviate potential errors created by improper parameterization. One potential solution to relying on expert opinion for elusive species

is to employ alternative statistical methods that do not require extensive parameterization of landscape genetic models.

Multivariate and regression methods (e.g., Legendre & Legendre 1998; Fortín & Legendre 2010; Wang 2013) could help alleviate challenges associated with parameterizing hypotheses about gene flow in elusive species. Multivariate and regression techniques offer considerable flexibility in the type of landscape and genetic variables that can be evaluated, negating reliance on *a priori* parameterization of landscape genetic hypotheses. Furthermore, multivariate and regression techniques are considered more robust than correlation statistics in genetics, and therefore, may limit erroneous conclusions due to type I or II errors. Recent studies have demonstrated the utility of multivariate and regression techniques for detecting complex, interacting influences on gene flow within continuously distributed, well-studied species, but have applied these techniques in a population-based framework (i.e., between genetic clusters; Reding *et al*. 2012 or between sampled areas; Blanchong *et al*. 2008; Robinson *et al*. 2012). Extending such methods to elusive species within a single study area or population requires individual-based approaches that can disentangle biologically relevant patterns in gene flow from spurious statistical correlations when there exists little biological data upon which to evaluate conclusions.

All individual-based landscape genetic statistics, even multivariate and regression techniques, are likely to suffer from some degree of error. Statistics aim to partition genetic variation according to specific landscape factors, but errors occur in individual-based studies because there is little genetic differentiation within a single population (i.e., little variance to partition). Errors are especially problematic in studies of elusive species

because statistical errors are difficult to quantify without corroborating field data and spatially biased sampling is almost unavoidable. Spatially biased sampling often causes misleading conclusions about gene flow (e.g., Schwartz & McKelvey 2010; Oyler-McCance *et al*. 2013), so studies in elusive species would benefit from explicitly testing the role of errors on landscape genetic statistics. Gene flow simulations offer a method to quantify errors in landscape genetic studies by providing a means of replication within a single landscape and control over processes that result in observed genetic variation (e.g., Epperson *et al*. 2010; Landguth *et al*. 2010). In particular, simulations can help separate false statistical significance (i.e., type I errors) from actual landscape effects on gene flow because multiple statistics in landscape genetics can have high type I error rates even in idealized sampling schemes (Balkenhol *et al*. 2009; Graves *et al*. 2013; Guillot & Rousset 2013). By quantifying type I errors using gene flow simulations, investigators can better understand how landscape heterogeneity impacts gene flow in elusive species, and separate those effects from sampling artifacts.

The American badger (*Taxidea taxus*) is one of the most elusive and poorly understood mesocarnivore species in North America. Badgers inhabit much of central and western North America, but much of their life history is unknown due to their nocturnal, semifossorial life style. Scarce telemetry studies illustrate dramatic differences in home range sizes and habitat selection among regions. For example, in British Columbia, badgers select habitats away from closed canopy forest (Apps *et al*. 2002), but those in Ohio used fallow fields and other uncultivated habitats within highly agricultural areas (Duquette & Gehrt 2014). Considerable individual variation in habitat selection also occurs within study areas, making it difficult to predict landscape effects on gene

flow. Based on the limited and likely variable life history data throughout most of North America and their elusive nature, this study on badgers suffered from both challenges associated with elusive species: limited life history data to parameterize landscape genetic models and unavoidable spatially-biased sampling.

In this study, we examined the utility of integrating multivariate and regression techniques with simulations for overcoming the challenges associated with limited life history and spatially biased sampling often observed in elusive species. To identify landscape factors that influence badger gene flow, this study used multivariate and regression techniques to test if isolation-by-distance (IBD; Wright 1943), badger's preference for treeless habitats, and a major topographic barrier within the study area are correlated with genetic variation of badgers in Wisconsin. Multivariate and regression techniques can test for these factors simultaneously, making them highly useful for disentangling multiple influences on gene flow without the need for parameterizing landscape hypotheses. Obtaining a large number of samples from the elusive badger required a statewide citizen-based effort, resulting in spatially biased sampling with samples clumped around populated areas, a factor that could lead to erroneous conclusions about gene flow. Therefore, we also performed gene flow simulations to calculate how often our chosen statistics falsely identified landscape effects (i.e., type I errors) to separate significant results created by sampling and actual landscape effects.

Methods

Study Area

Our study area (110,745 km^2) encompasses the state of Wisconsin in the Upper Midwest, United States. The landscape exhibits a transition from a mixture of native grasslands and agriculture in the south to more forested habitats as latitude increases. Badger activity has been recorded in every county in Wisconsin based on citizen-based monitoring of badgers from 2009-2014 and Wisconsin Department of Natural Resources mammal surveys from 1987-2008 (Wydeven *et al*. 1998; Kitchell 2008); both suggest badgers have a relatively continuous distribution throughout Wisconsin. Despite their continuous distribution within Wisconsin, recent genetic evidence suggests that badgers in Wisconsin represent a unique genetic population within North America due to the Mississippi River and Great Lakes (Chapter 3).

Sample Collection

Since badgers are protected from harvest in Wisconsin as a Species with Information Needs (Wisconsin Department of Natural Resources, 2008), 250 individuals were sampled via road-kills (n = 136 tissue samples) and non-invasive hair collections (n = 114 hair samples) from 2005-2013. Hair collection involved attaching a snare (modified from British Columbia Ministry of Environment Ecosystems Branch for the Resources Information Standards Committee 2007) to the entrance of an active burrow and waiting overnight for the animal to pass under the snare. Snares were designed to line the entrance of the burrow, so hair was typically pulled from both the back and sides of an animal. A successful hair collection typically contained 20 to 50 hairs with intact roots, ensuring a sufficient number of hairs for molecular analysis. No burrows were

sampled repeatedly, and we limited hairs to long banded hairs that are typical of adults to help prevent sampling of younger animals (juveniles and kits).

Location information for each individual was recorded in latitude and longitude coordinates. For a few individuals with less precise location information (i.e., Counties, Public Land Survey System locations; n = 42), we utilized the "Create Random Points" function in ArcMap v. 10.1 to assign a spatial coordinate. In Wisconsin, the maximum amount of ambiguity around a point was approximately 10 km (locations as counties). No change in results was detected by excluding individuals with spatial ambiguity from analysis.

Laboratory Methods

DNA was extracted using Qiagen DNEasy blood and tissue kits for both tissues and hairs. Extractions used a 1 mm^3 piece of tissue or 10 to 20 hairs with intact follicles in a user-refined extraction protocol (Qiagen 2006). Hairs were processed on different days than tissue samples in dedicated laboratory space. We amplified all samples at 12 microsatellite loci (Table S2.1) developed in American badger (Tt-1, Tt-2, Tt-3, and Tt-4; Davis & Strobeck 1998), American mink (*Neovison vison*; Mvis072; Fleming *et al*. 2002), American marten (*Martes americana*; Ma-1; Davis & Strobeck 1998), and European badger (*Meles meles*; Mel112, Mel101, Mel111, Mel108, Mel14, and Mel1: Carpenter *et al*. 2003; Domingo-Roura *et al*. 2003). Multiplex polymerase chain reactions (PCRs) were conducted in sets of 2-4 primers in 10 uL reaction volumes (Table S2.1). Amplified products were genotyped on an ABI 3730 DNA Analyzer at the

University of Wisconsin Biotechnology Center, and alleles were sized using the program GeneMarker (SoftGenetics LLC, State College, PA, USA).

We utilized a comparative multi-tube approach for genotyping hair samples (Frantz *et al*. 2003; modified from Navidi *et al*. 1992 and Taberlet *et al*. 1996) where each hair extract (two for most samples; n = 85, single extraction; n = 39) was genotyped three (heterozygotes) to seven times (homozygotes) to validate resultant genotypes. Tissue samples were collected from road-killed animals and varied considerably in quality, so we also re-extracted and re-genotyped 50% of tissue samples (n = 68). All resultant homozygotes and 25% of heterozygotes in tissues were re-genotyped as a final check. In total, we identified 6 instances of allelic dropout (all hairs) in the 918 repeated genotypes (0.65% error rate). Any individuals with fewer than 10 genotypes were then culled from the dataset (17 hair samples culled; 233 retained individuals).

Statistical Analysis

We used two complementary Bayesian clustering approaches (non-spatial STRUCTURE 2.2.3; Pritchard *et al*. 2000 and spatial BAPS 5; Corander *et al*. 2008) to characterize population structure within Wisconsin. In our non-spatial approach, we employed Bayesian clustering in STRUCTURE with five independent runs (100,000 MCMC burn-in, 100,000 permutations) at each hypothesized number of genetic clusters (K) under the admixture, correlated alleles model (Pritchard *et al*. 2000). The optimal value for K among tested values (K = 1-10) was determined using Evanno *et al*.'s (2005) ΔK estimator because likelihood values plateaued and variances among runs grew larger at values of K above the optimum (Pritchard *et al*. 2000; Figure S2.1). Once the optimal

K was identified, five longer runs of 1,000,000 MCMC burn-in, 1,000,000 permutations were conducted to calculate the proportion of each individual's genome that belongs to each cluster (*q*). Average *q*-values across the 5 runs were calculated in CLUMPP (Jakobsson & Rosenberg 2007), and individuals were assigned to a cluster based on their highest *q*.

In our spatially informed approach, we employed BAPS 5. We tested $K = 1$ through 10 (5 replicates per *K*) using the "Spatial Clustering of Individuals" option in BAPS. Maximum likelihood and highest posterior probability were used to determine the optimal number of genetic clusters in the sample. Admixture between inferred clusters was calculated using 500 simulations based on observed allele frequencies.

For the total sample and the inferred clusters, we calculated population-specific measures of genetic diversity (allelic richness, number of alleles, and heterozygosity, and F_{IS}) and genetic differentiation among clusters (F_{ST}) in the R package diveRsity (Keenan 2013). Deviations from Hardy-Weinberg (HWE) and Linkage (LE) equilibria were calculated in GENEPOP (Raymond & Rousset 1995) using a corrected alpha for multiple tests ($\alpha = 0.012$; false discovery rate; Benjamini & Yekutieli 2001).

To test for the effect of geographic distance on genetic differentiation, we quantified patterns of IBD within Wisconsin using two complementary approaches. We used simple Mantel tests to test for an association between matrices of pairwise genetic and geographic distances and dbRDA to test for significant effects of geography (latitude and longitude) on the distribution of genetic variation across the study area. Both Mantel tests and dbRDAs were conducted in the R package VEGAN using the functions "mantel" and "capscale" (Oksanen *et al*. 2008).

In addition to tests for IBD, spatial autocorrelations were used to detect departures from random mating (i.e., panmixia) within 5 km distance categories. Individuals separated by small geographic distances are expected to exhibit positive spatial autocorrelations (i.e., be more genetically similar than expected under panmixia). Statistical significance in spatial autocorrelations was assessed after 1000 permutations using custom code in R. Mantel tests, dbRDA, and spatial autocorrelations were run with the heavily sampled counties subsampled by factors of 5 (0, 5, 10, 15, and 20 individuals were included from Bayfield, Dane, and Iowa Counties) to alleviate any potential sampling bias.

Population structure is often influenced by discrete barriers (isolation-by-barrier; IBB) in addition to geographic distance (IBD). We used a partial Mantel test to determine correlational significance between a genetic distance matrix (Rousset's *a*; Rousset 2000) and a barrier matrix while controlling for geographic distance. The barrier matrix was a binary indicator of whether a pair of individuals was on the same (0) or different (1) sides of the Wisconsin River, the most prominent potential barrier to badger gene flow in Wisconsin. Calculations were performed using the R package VEGAN, and statistical significance was assessed via Pearson correlation coefficients after 1000 permutations.

In nature, many different landscape factors can work in tandem to create observed patterns of genetic variation (e.g., Cushman *et al*. 2006). Thus, we sought to incorporate both barriers and ecological variables (collectively called landscape factors) into a cumulative model that explains how all landscape factors influence genetic variation in badgers. Ecological variables included level III ecoregions and land cover (Figure 2.1).

We used level III Ecoregions to define broad scale regions of similar abiotic (i.e., climate and soil) and biotic assemblages (in Wisconsin: Northern Lakes and Forest; North Central Hardwood Forest, Driftless Area, and Southeastern Wisconsin Till; Omernik 1987). Land cover variables were produced via an intersection of two rasters: native soil associations within Wisconsin (Hole 1976) and land cover (NLCD2006; Fry *et al*. 2011), and then summarized into three land cover categories (Native Open, Forest, and Agriculture) that represent a continuum of habitat suitability for badger. Proportions of each land cover category were calculated within 5 km circular buffers drawn around each badger location. Proportions of Native Open, Forest, and Agriculture were highly correlated ($r = -0.35$ to -0.65, all $p < 0.001$), so only one land cover variable was used within each statistical model at a time.

Multivariate techniques like spatial principal components analysis (sPCA; Jombart *et al*. 2008) are highly effective at detecting both discrete barriers and genetic gradients, which makes them ideal for disentangling complex patterns of gene flow (Jombart *et al*. 2008). We used the R package ADEGENET (Jombart 2008) to perform sPCA calculations and significance tests for global and local patterns using 999 permutations. We used an inverse distance-weighting network so that all badgers were considered neighbors. The first two sPCA axes explained the most variation (0.0889 and 0.0571, all others < 0.0459; Figure S2.3), and were retained as dependent variables in landscape genetic analysis (Figure 2.2).

We utilized two approaches to distinguish between landscape factors and geographic distance, both of which can drive patterns in sPCA axes (Jombart *et al*. 2008). We performed a partial redundancy analysis (RDA), a constrained ordination technique

that is the multivariate analog to simple linear regression (Legendre & Legendre 1998). The two retained sPCA axes were analyzed together as explanatory variables with both latitude and longitude in the conditional matrix (i.e., variables to be controlled for in final models). The barrier (Wisconsin River) and ecological (Ecoregion or Agriculture) variables were included as predictors. We performed a stepwise model selection procedure to identify the variables that best explained genetic differences among individuals. We used the R package VEGAN, with significance assessed with 1000 permutations using function "ordistep" (Oksanen *et al*. 2008).

We used spatially lagged regression models as a second approach to evaluate the evidence for landscape influences on sPCA axes by controlling for the potentially confounding influence of IBD. Spatially lagged regression models evaluate each sPCA axis individually, which provides further evidence for specific influences on patterns found within the sPCA. Spatially lagged regression models control for IBD by including a spatial weighting matrix (W_{ij}) and ρ, a parameter that accounts for the lack of independence between individuals (Legendre & Legendre 1998). The weighting matrix was constructed based on an inverse weighting calculation as this procedure is thought to best approximate spatial autocorrelation under IBD (Robinson *et al*. 2012). Regression models: $G_i = \rho * W_{ij} * G_j + \beta * X + \varepsilon$ included sPCA scores for a focal individual *i* (G_i) and all other individuals (G_j) along with explanatory barrier and landscape variables (X), their estimated effects (β), and residual error (ε). A LaGrange Multiplier test then tests for spatial autocorrelation within residuals of the regressions where a significant test indicates autocorrelation remains within the data. One major advantage of spatial regression is the ability to conduct model selection procedures such as Aikaike

Information Criterion (AIC). Model selection followed Burnham & Anderson's (2002) method where models with a $\Delta AIC < 2.0$ were considered candidate models for explaining a sPCA axis. Parameter estimates were produced through model averaging of all candidate models where significant parameters do not include zero (Burnham & Anderson 2002). All spatial regression methods were performed using the "lagsarlm" function in R package SPDEP (Bivand *et al*. 2011).

To assess type I errors within sPCA, partial RDA, and spatially lagged regression models, we simulated 100 populations where only geographic distance influenced gene flow in the program CDPOP v. 1.4 (Landguth & Cushman 2009). Populations were simulated using an inverse square dispersal distribution because strong sex-biased dispersal has not been recorded in badgers (Messick & Hornocker 1981; Hoodicoff 2003; Kierepka *et al*. 2012). Mating was sexual with replacement and generations were overlapping. The initial population consisted of 1,400 individuals; 1,167 were placed randomly on Wisconsin's landscape excluding urban and open water habitats and 233 were placed at geographic coordinates identical to our empirical dataset. After 250 generations, we sampled the 233 individuals that had the same geographic locations as the empirical dataset. We performed sPCA, partial RDA, and spatially lagged regression analyses for each simulated population, defining type I errors as false detection of a barrier (Wisconsin River) or ecological (Ecoregion, land cover) factor as a driver of population genetic structure in more than 5 of the 100 simulated populations (alpha = 0.05).

Results

In the BAPS analysis, the optimal solution was K = 1 (-11197.19). The STRUCTURE analysis indicated stronger support for K = 2 than for K = 1 (Figure S2.1a). However, an inspection of the STRUCTURE assignments revealed a gradient pattern (Figure S2.1b). The resultant gradient in q-values and results from BAPS suggests that the inferred structure was likely an artifact of IBD or spatial autocorrelations because both can create false genetic clusters in Bayesian programs (Frantz *et al*. 2009; Schwartz & McKelvey 2009).

We observed a heterozygote deficiency within Wisconsin over all loci in a global analysis (Table 2.1) and for seven individual loci (all $p < 0.001$). These deviations from HWE are expected if any deviation from panmixia exists within the dataset. All loci were highly polymorphic ranging from 7 to 15 alleles per locus (average = 11.58 alleles/locus), and showed no linkage disequilibrium (all $p > 0.025$).

Results from the dbRDA analysis supported the Bayesian clustering and HWE analyses that indicated a role for geography driving gene flow patterns. However, the dbRDA analysis was significant for only latitude ($F = 2.404$, $p = 0.008$), not longitude ($F = 1.009$, $p = 0.124$). Also, the simple Mantel test did not show a significant correlation between matrices of genetic and geographic distances ($r = 0.012$, $p = 0.0312$). These results suggest that variability in longitude across Wisconsin may be too small to detect IBD.

Despite the equivocal evidence for IBD in dbRDA and Mantel test, we found support for fine-scale spatial autocorrelations. Genetic distances between proximate individuals (25 km or less) were smaller than expected under panmixia in the spatial

autocorrelation (Figure S2.2). In particular, individuals separated by 5 km were highly autocorrelated indicating that individuals found closer together were more genetically similar. One explanation for the strong positive autocorrelations at 5 km is that we sampled relatives in Dane, Iowa, and Bayfield Counties where 59/90 pair-wise comparisons under 5 km occurred. However, most of these samples were road-killed individuals sampled across years or live captured adult animals at separate burrows. We completely removed these counties in another set of spatial autocorrelation tests, and the strong positive autocorrelation remained at 5 km. Therefore, we argue that the positive spatial autocorrelation is not an artifact of sampling relatives.

We found no evidence for IBB resulting from the Wisconsin River in either Bayesian analyses or partial Mantel tests. Visual inspection of population assignments from STRUCTURE revealed a gradient in q-values that lacked a strong genetic break along the Wisconsin River. The partial Mantel test also revealed a lack of barrier effect ($r = -0.0266$, $p = 0.992$).

Spatial PCA axes revealed two main patterns within Wisconsin: a latitudinal cline across the entire state (Axis 1) and an area of high genetic similarity in central Wisconsin (Axis 2; Figure 2.2). Both axes had signatures of spatial autocorrelation (Moran's $I = 0.572$ and 0.301), and Axis 1 explained more variation in genetic diversity than Axis 2 (variance = 0.175 vs. 0.152). Loci Mel1, Tt-2, Mvis072, Mel101, and Mel108 were most informative for Axis 1 and Mel111, Tt-2, and Mel1 were most useful for Axis 2. The Monte-Carlo test confirmed the existence of at least one global pattern (observed: 0.010, $p = 0.001$) but no local pattern (observed: 0.006, $p = 0.362$).

Results from the partial RDA and spatially lagged regressions indicated that the Wisconsin River and Agriculture were correlated with the two sPCA axes. In the partial RDA analysis, both the Wisconsin River ($F = 22.44$, $p < 0.001$) and Agriculture ($F = 59.416$, $p < 0.001$) were retained after model selection. Ecoregions, in contrast, were not included in the final RDA models. Only Agriculture was found to be associated with sPCA spatially lagged scores for both sPCA axes; Forest or Native Open were not retained as significant variables when used as the land cover variables.

All top spatially lagged regression models for Axis 1 contained the Wisconsin River and model averaging revealed that the parameter estimate did not include zero (parameter estimate = -0.225 ± 0.144; Table 2.2). However, observed models also had relatively high residual spatial autocorrelations (LaGrange Multiplier test: = 1.967-4.768, $p = 0.0290$-0.116), suggesting that IBD was also associated with Axis 1. All top models for Axis 2 included Agriculture (Agriculture, Agriculture + River, Agriculture + Ecoregion; Table 2.2). Model averaging between these three models indicated that Agriculture was the only significant variable (model averaged parameter: -0.318 ± 0.284). Observed models indicated a lack of remaining spatial autocorrelation in Axis 2 (LaGrange test: 0.379-0.517, $p = 0.439$-0.538). Like the RDA analysis, when Forest and Native Open were included as land cover variables, they were not significantly associated with sPCA spatially lagged scores of either axis.

Simulations revealed similar type I error rates for partial RDA and spatially lagged regression (Figure 2.3). The Wisconsin River (52-57/100 populations) and Ecoregions (68-76/100 populations) were falsely detected as ecological variables affecting gene flow much more frequently than our 5 tests out of 100 cut-off. A likely

cause of high error rates was that residual autocorrelation (i.e., IBD) remained within the data (LaGrange Multiplier test: all $p < 0.062$) in almost all cases of significance. This high false significance rate in simulated populations suggests that our finding of an association between the Wisconsin River and spatially lagged scores from Axis 1 in our empirical dataset was likely an error. A lack of barrier effect for the Wisconsin River was also supported by the Bayesian analyses and partial Mantel test. In contrast, Agriculture was falsely detected as an ecological variable affecting gene flow in less than 5% of populations regardless of the test (3-5/100 populations; Figure 2.3). Therefore, our finding of an association between Agriculture and the spatially lagged regression scores was unlikely to be a type I error in our empirical dataset, and represents an ecological variable that likely impacts gene flow.

Discussion

There are many challenges associated with elusive species that can affect conclusions about how features of the landscape influence gene flow. In badgers, both limited life history data and spatially biased sampling were obstacles that could potentially confound our interpretation of landscape genetic statistics. Our study demonstrates that combining multivariate and regression statistics with simulations to quantify errors can help disentangle potential errors from landscape effects on gene flow in an individual-based framework. RDA and spatially lagged regression did not require parameterized hypotheses, which was critical for this study given the nearly complete lack of relevant life history data. With simulations, we were able to quantify the

confounding effects of spatially biased sampling to separate type I errors from biologically relevant landscape effects on gene flow. Following error assessment, our genetic dataset indicated that geographic distance is the strongest influence on badger gene flow within Wisconsin, with agriculture playing a lesser role.

Geographic distance was the primary driver of gene flow in badgers, as evidenced by dbRDA and spatial autocorrelations. Both axes in the sPCA had strong signatures of spatial autocorrelation, particularly in axis 1, further demonstrating that geographic distance influenced gene flow in badgers. The non-significant simple Mantel test and dbRDA for longitude were inconsistent with the sPCA results, which suggests that geographic distance is only important at local scales. Positive spatial autocorrelations were particularly pronounced under 5 km in our dataset, which could reflect either a behavioral mechanism or sampling artifact. Restricted dispersal can occur in mammals due to philopatry, particularly in females (Greenwood 1980), but little evidence has been found for philopatry in American badgers (Messick & Hornocker 1981; Kierepka *et al*. 2012). However, dispersal regimes can vary according to habitat quality where dispersal is more restricted in suitable habitat (e.g., Broquet *et al*. 2006; Frantz *et al*. 2009), so high spatial autocorrelations may indicate badgers may exhibit some degree of restricted dispersal in Wisconsin. Badgers were highly genetically similar in areas with more suitable habitat, but these areas were also the most heavily sampled. Within these heavily sampled counties, proximate pairs of individuals were often collected as road-killed animals, making it difficult to determine whether individuals were killed within suitable habitat or during dispersal though matrix habitat. Regardless of the mechanism, positive spatial autocorrelations at fine scales is a fairly ubiquitous factor influencing spatial

patterns of genetic variation in other highly mobile carnivores (e.g., Schwartz *et al*. 2006; Cegelski *et al*. 2006; Zalewski *et al*. 2009; Croteau *et al*. 2010), so it is not surprising to find that geographic distance exhibits a strong influence on gene flow in badgers as well.

After controlling for the influence of geographic distance across Wisconsin, our multivariate and regression statistics detected the Wisconsin River as a potential influence on patterns in sPCA axis 1. However, considerable residual autocorrelation remained in the spatially lagged regressions and differentiation in sPCA axis 1 appeared to be greatest between the southeastern and northwestern areas of Wisconsin (i.e., consistent with geographic distance). Simulated populations in which geographic distance was the only influence on gene flow also frequently detected a barrier effect of the Wisconsin River, indicating that the statistically significant effect we observed in our dataset is likely a sampling artifact. Badger activity in the largely forested northcentral areas of Wisconsin was rarely reported, so most sampled individuals coded as west of the Wisconsin River were from northwestern Wisconsin. Also, the Wisconsin River occurs in the center of our study area, so removing the effect of geography in the spatial regression and RDA was difficult as evidenced by the high residual autocorrelation. Therefore, the gap in sampling west of the Wisconsin River combined with the underlying isolating effects of geographic distance appear sufficient to create the statistically significant associations between patterns of genetic variation and the Wisconsin River.

Unlike the Wisconsin River, type I error rates were low for Agriculture, which suggests agricultural landscapes influence gene flow within Wisconsin badgers. In this case, agriculture appeared to facilitate gene flow more readily than other habitats as

evidenced by the high genetic similarity of individuals (Axis 2 of sPCA) within agricultural habitats. Optimal habitats often facilitate gene flow (Cushman *et al*. 2006; Schwartz *et al*. 2009), but badgers generally avoid agriculture in other portions of their range (Messick & Hornocker 1981; Warner & Ver Steeg 1995; Duquette & Gehrt 2014). Avoidance behavior suggests that agriculture is not an optimal habitat for badgers, so understanding the exact role of agriculture in driving gene flow across Wisconsin is not straightforward without corroborating field data on dispersal.

Several mechanisms could explain why agricultural habitat is correlated with genetic variation in badgers. First, agriculture could be significant because of its tight correlation with Native Open, the likely preferred badger habitat. Native Open was not significantly associated with either sPCA axis, but sampling within native open habitats was relatively sparse compared to agriculture. Robust sampling occurred within the northwestern native open habitats unlike the native open habitats within the largely agricultural southern Wisconsin. In southern Wisconsin, native open habitats were often interspersed with agriculture, but spatial data may have lacked the resolution to capture fine-scale habitats (e.g., native open habitats along fencerows) necessary for dispersal (Anderson *et al*. 2010). Badgers in other agricultural landscapes utilized linear corridors along agricultural areas for movement (Duquette & Gehrt 2014), so agriculture as a whole may appear to facilitate gene flow in Wisconsin. Alternatively, density-dependent dispersal could promote gene flow from suitable habitats into agriculture (e.g., Carr *et al*. 2007). As resident badgers tend to remain in the same area over consecutive years (Lindzey 1978; Messick & Hornocker 1981), a lack of unoccupied areas within suitable habitats (i.e., native open) may force transient and juvenile individuals into agriculture, a

sub-optimal habitat. Both of these putative mechanisms would produce the pattern observed in sPCA Axis 2 where individuals in high intensity agricultural habitats show high genetic similarity to those in surrounding suitable habitats (e.g., Native Open or pasturelands in central and southwestern Wisconsin, respectively). Both mechanisms also would produce elevated F_{IS} values in badgers in agricultural habitat relative to other habitats, a pattern that was present in our study but not significant. Determining the exact role of agriculture in badger gene flow is difficult because we do not know whether animals sampled in agriculture were transient dispersers or had established home ranges. Despite an inability to distinguish between potential mechanisms causing landscape genetic patterns in badgers, genetic data can be used to develop focused hypotheses (i.e., how badgers disperse in agricultural habitats) that with targeted sampling, can test how these specific landscape features impact connectivity.

When interpreted properly through rigorous error testing, regression and multivariate statistics are well suited to elusive species like badgers in individual-based landscape genetics. One benefit of multivariate and regression techniques is that they offer considerable flexibility in both landscape and genetic variables. To date, population-based studies have successfully incorporated a wide array of variables into multivariate and regression techniques (e.g., Balkenhol *et al*. 2009; Reding *et al*. 2012; Robinson *et al*. 2012), but many of those population-specific variables are not applicable to individual-based approaches. Our study with badgers emphasizes that multivariate and regression techniques are equally useful in individual-based studies with appropriate choices in landscape and genetic variables. We selected our categorical land cover variables based on the limited telemetry studies in badgers (e.g., Apps *et al*. 2002;

Duquette & Gehrt 2014) because no movement or habitat preference data is available in Wisconsin or nearby areas. Deriving pair-wise connectivity metrics for badgers would have had considerable uncertainty given the substantial individual and population variation in demography and habitat associations throughout their range (Messick & Hornocker 1981; Apps *et al*. 2002; Duquette & Gehrt 2014). Even in well-studied species, landscape genetic patterns vary between landscapes, primarily due to the presence or spatial arrangement of important factors (e.g., Short Bull *et al*. 2011). RDA and spatially lagged regressions do not necessitate extrapolation to produce pair-wise connectivity metrics, and can still detect subtle impacts on gene flow as seen with agriculture in badgers. Therefore, their flexibility in variable type makes multivariate and regression techniques attractive for a myriad of landscape genetic studies, including those focused on species with limited life history information.

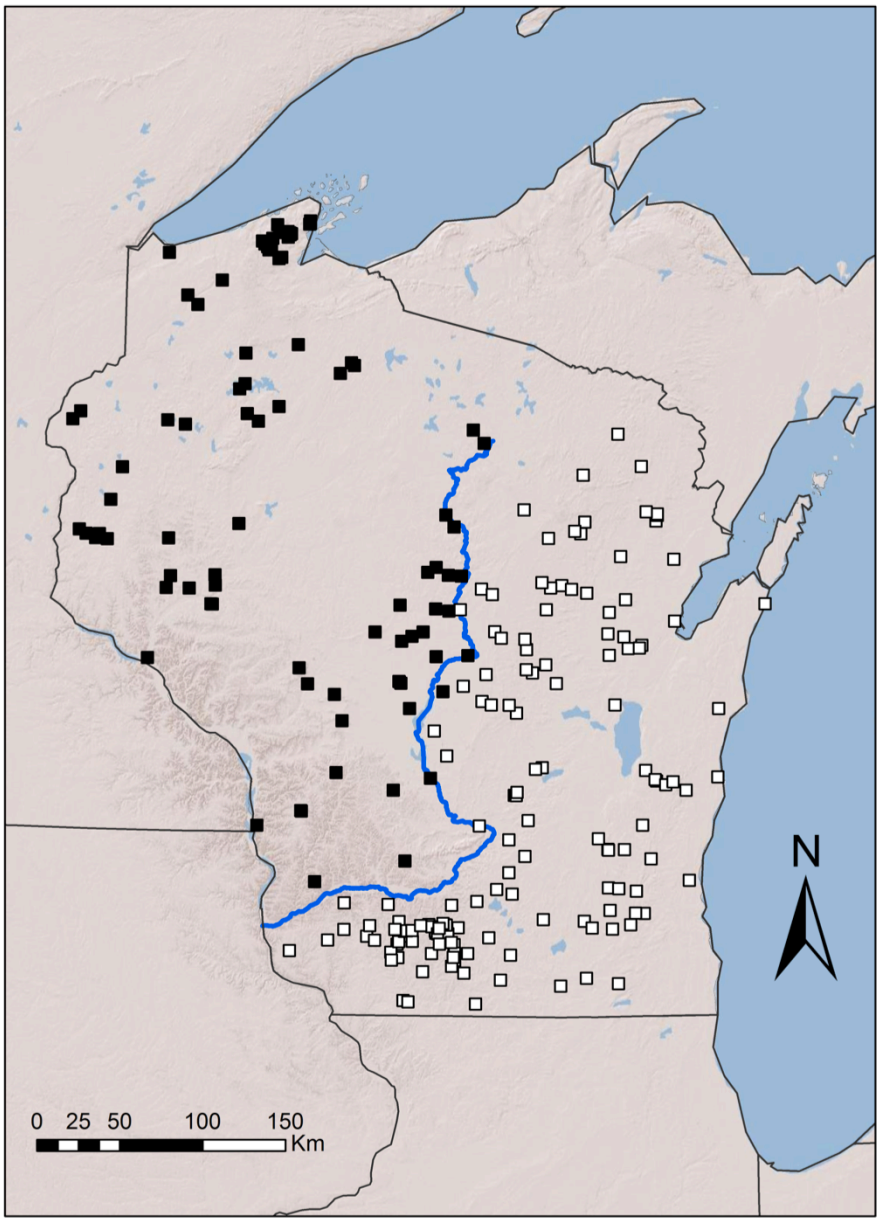
In addition to flexibility, multivariate and regression techniques may be particularly useful in conservation because both techniques have high power to detect fine-scale landscape genetic patterns (Balkenhol *et al*. 2009; Fortín & Legendre 2010). In this study, both RDA and spatially lagged regressions detected agriculture's subtle impact on gene flow despite the strong influence of geographic distance on genetic variation. Based on their ability to detect landscape genetic patterns in simulated spatial gradients with considerable noise (Fortín & Legendre 2010), multivariate and regression techniques likely would perform well in situations with complex patterns in gene flow (e.g., Bowen *et al*. 2005; Kamler *et al*. 2013) and biased sampling. Although multivariate and regression techniques are flexible, statistically robust, and provide a means of multi-model inference, error assessment is also critical as all statistics utilized in this study

were also vulnerable to type I errors. When combined with simulations to quantify potential errors, multivariate and regression techniques can be a powerful tool to inform conservation and management efforts in poorly understood species.

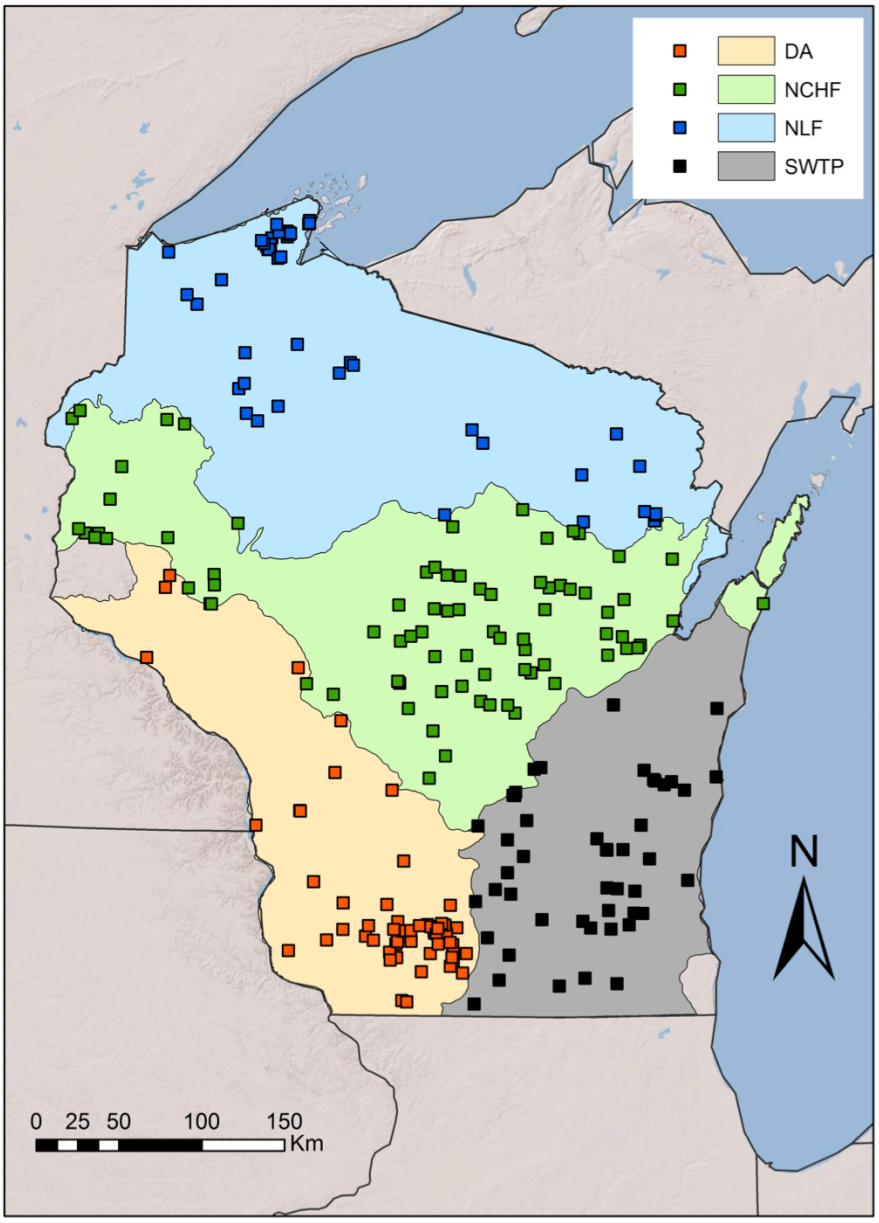
Overall, this study provides important information for future management of the American badger in Wisconsin, a population that is genetically distinct due to the Mississippi River and Great Lakes (Chapter 3). Badger gene flow within Wisconsin is largely unrestricted despite heterogeneous habitat composition and the presence of large riverine barriers. Future studies would be helpful to assess potential mechanisms that could explain the relationship between agriculture and genetic variation observed in this study (i.e., conduits of dispersal or density dependent dispersal). This study demonstrates the utility of multivariate and regression methods within individual-based landscape genetics, which to date, were largely restricted to population-based investigations of well-studied species. With multivariate and regression methods and explicit error assessment, individual-based landscape genetic approaches can provide valuable insights into how landscape heterogeneity impacts genetic variation despite limited life history and biased sampling.

Figure 2.1. Predictor variables for each georeferenced badger describe its location according to three landscape features (Wisconsin River, level III Ecoregion, and land cover). For the Wisconsin River (blue line), each individual was coded as either east or west of the river (a). All badger locations fell within one of four level III Ecoregions within Wisconsin (b): Driftless Area (DA), Northcentral Hardwood Forests (NCHF), Northern Lakes and Forest (NLF), and Southwestern Wisconsin Till Plain (SWTP). Land cover data for each badger was calculated as percent land cover within a circular buffer surrounding each badger's location (most predominant land cover class given by point's color; c).

(a)



(b)



(c)

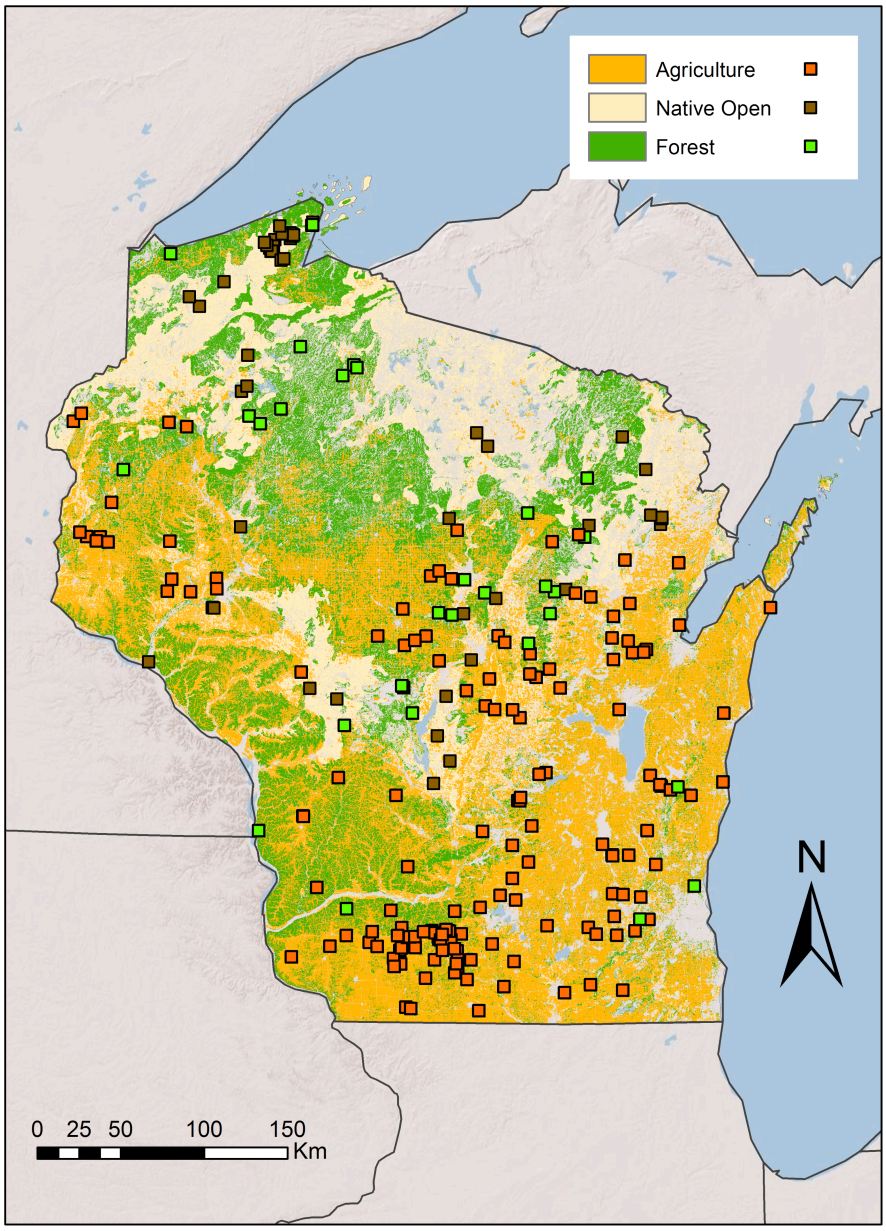
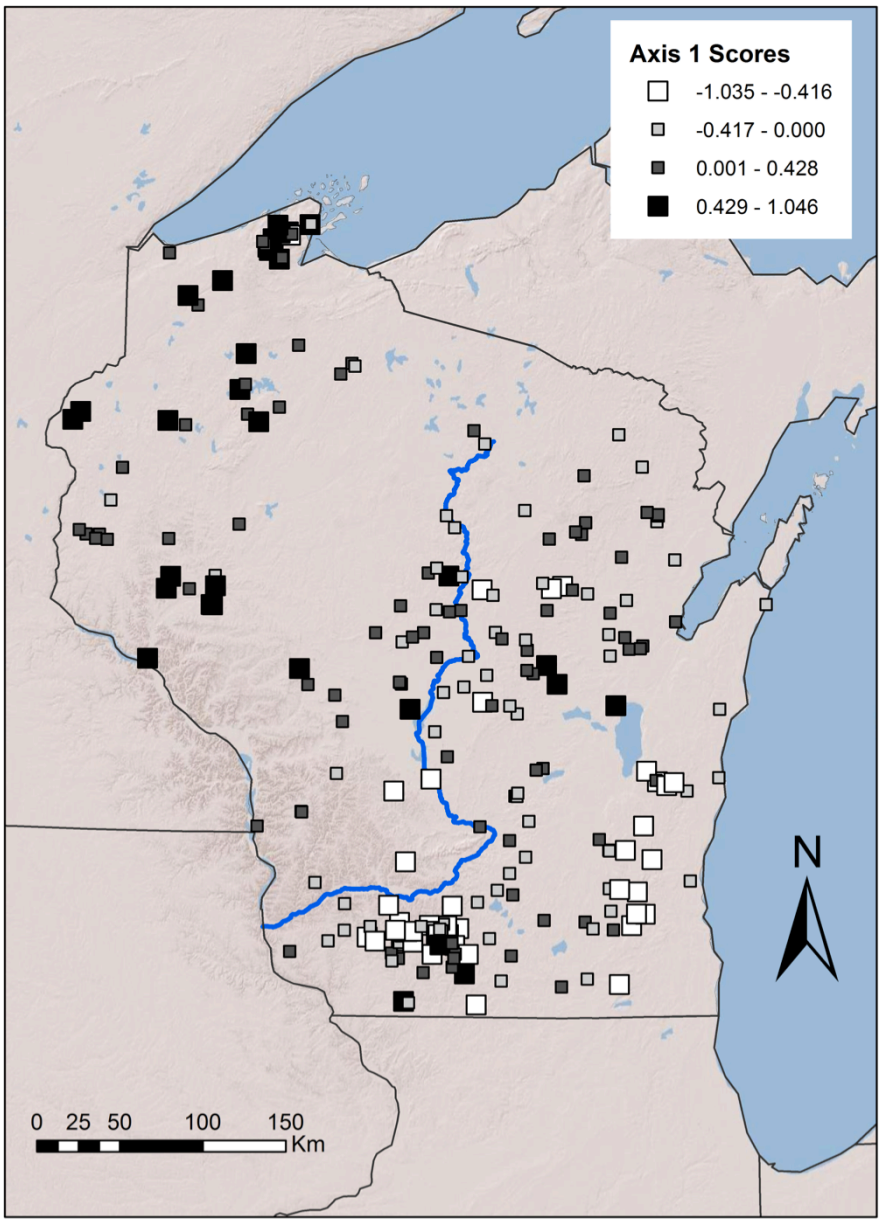


Figure 2.2. Spatially lagged scores for the first two sPCA axes for Wisconsin badgers. Scores from Axis 1 (a) and 2 (b) were correlated with the Wisconsin River and percent agriculture respectively. Dark colors (black and dark grey) represent negative sPCA scores while positive values are light in color (white and light grey). More extreme values in sPCA axes are displayed with larger squares.

(a)



(b)

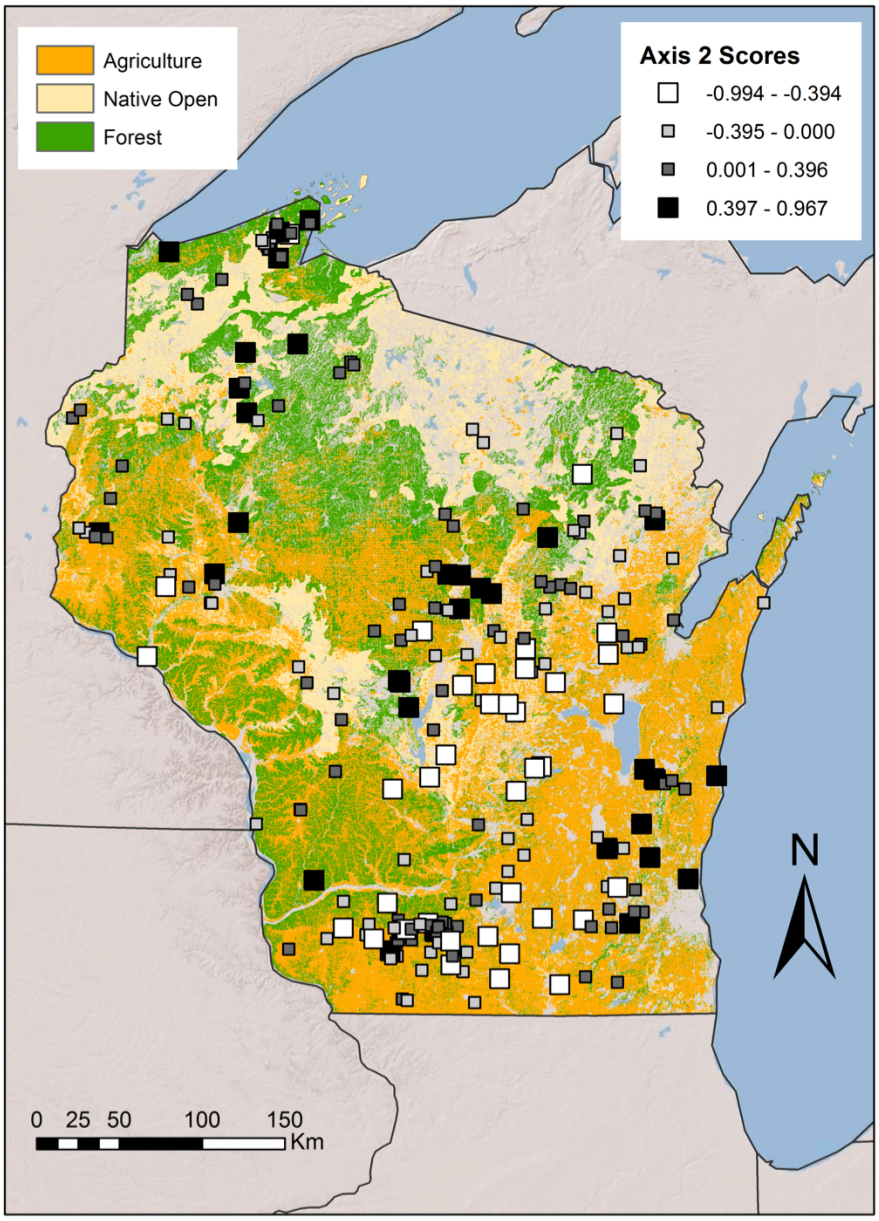

Agriculture
Native Open
Forest
Axis 2 Scores
-0.994 - -0.394
-0.395 - 0.000
0.001 - 0.396
0.397 - 0.967
N
0 25 50 100 150
Km

Figure 2.3. Percent of false significant tests (type I errors) for the Wisconsin River, level III Ecoregions, and Agriculture calculated in 100 simulated IBD populations. Both partial RDAs and spatially lagged regressions (sReg) for each axis separately incorrectly identified the Wisconsin River and level III Ecoregions as influences on gene flow more often than Agriculture.

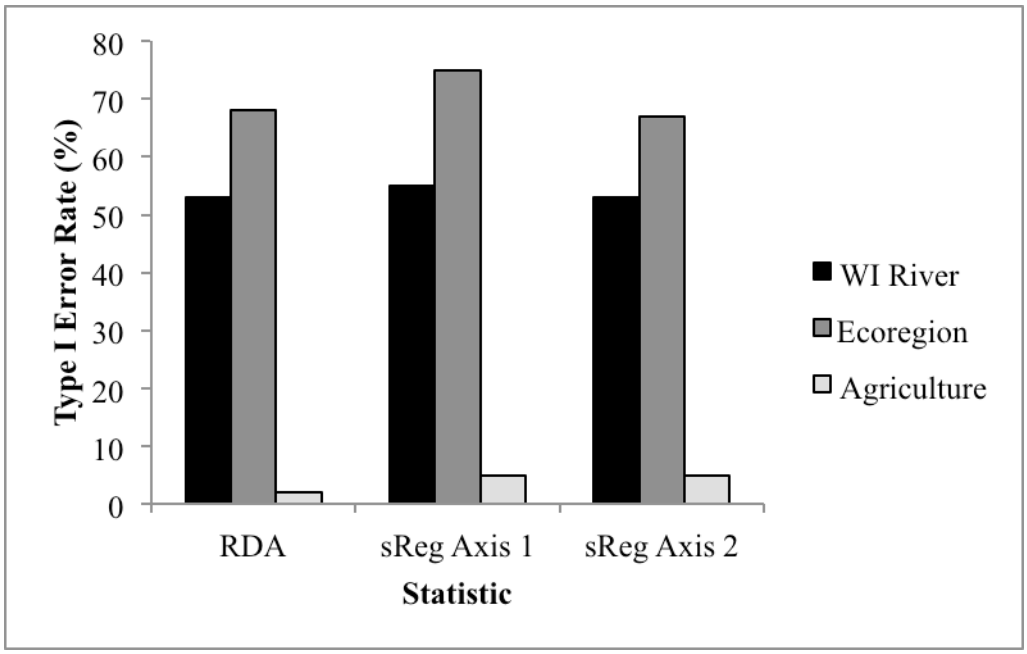
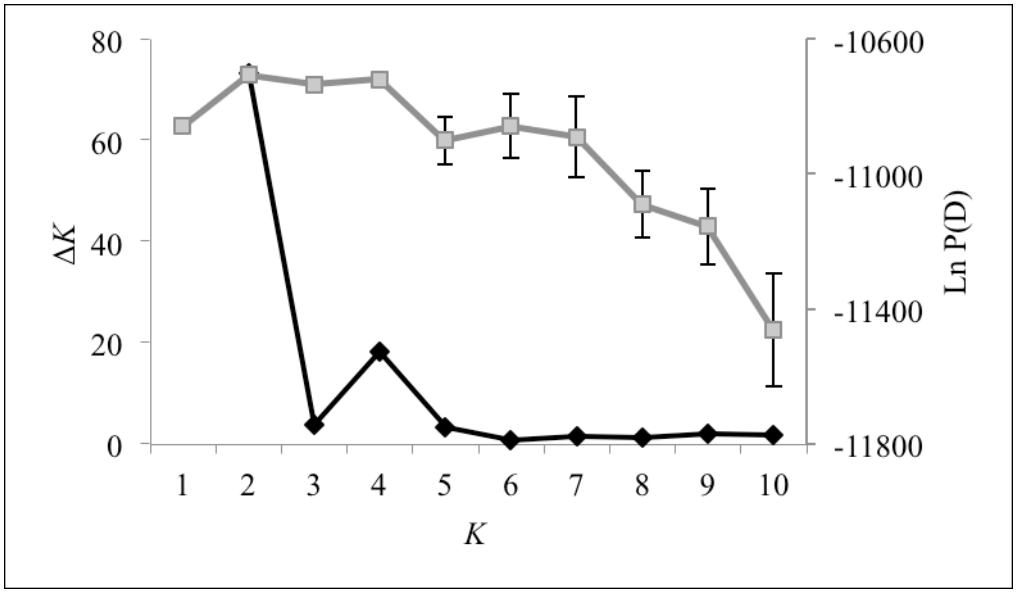


Figure S2.1. Results of Structure analyses from $K = 1$ to $K = 10$. Maximum likelihoods (grey squares) and ΔK values (black diamonds) occurred at $K = 2$ (a). Error bars correspond to standard deviations around each likelihood across five runs per K. A heat map was constructed from averaged q-values for $K = 2$ based on inverse distance weighting (b). The heat map revealed no conspicuous barriers, but rather exhibited a gradient from northwest to southeast.

(a)



(b)

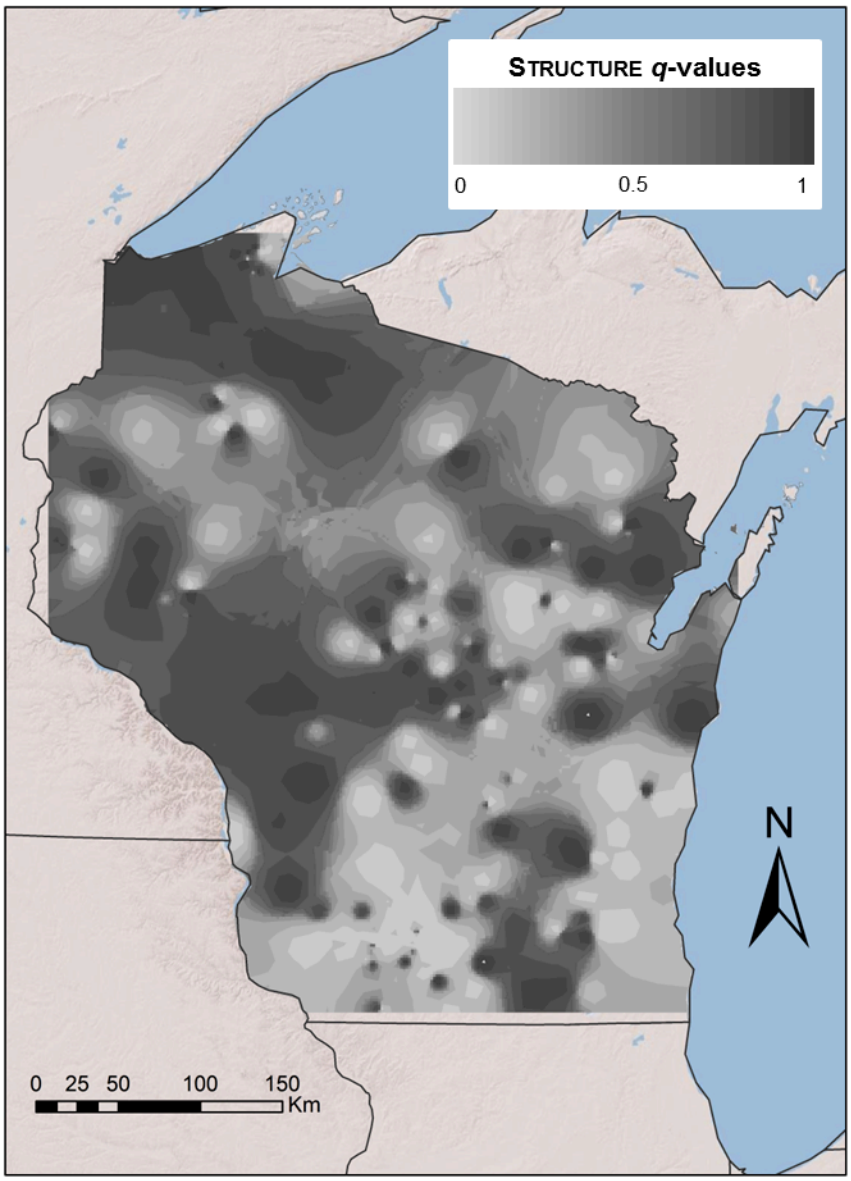


Figure S2.2. Spatial autocorrelation for all badgers within Wisconsin. Dotted lines correspond to 95% confidence intervals derived from 1000 permutations of Rousset's *a* within each 5 km distance class. Each point represents average Rousset's *a* at a distance class, and those below the 95% confidence intervals (< 20 km) had lower genetic distance than expected by random mating.

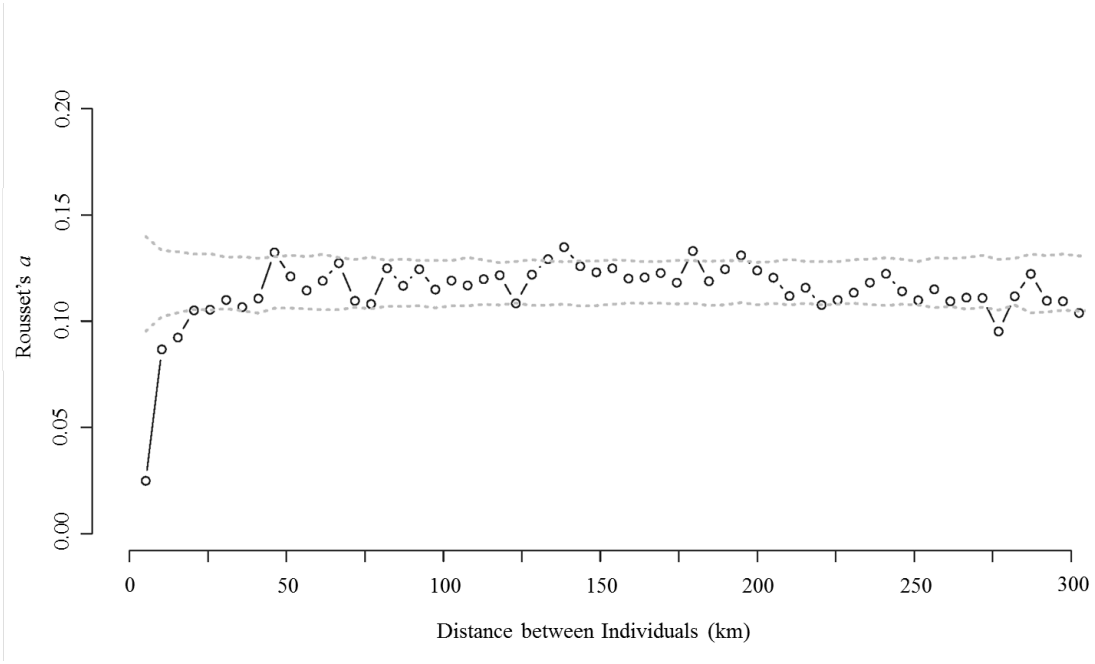
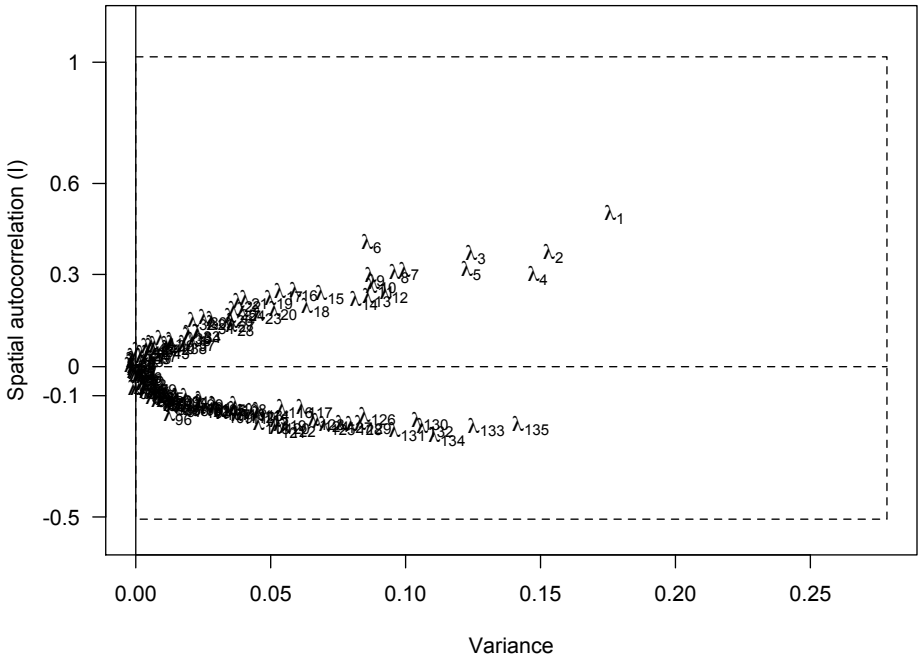


Figure S2.3. Eigenvalues (a) and screeplot of sPCA (b) displaying the variance explained and spatial autocorrelation within each sPCA axis. Axis 1 and 2 were retained for further analysis because they had the highest eigenvalues and explained the most variance within the dataset.

(a)



(b)

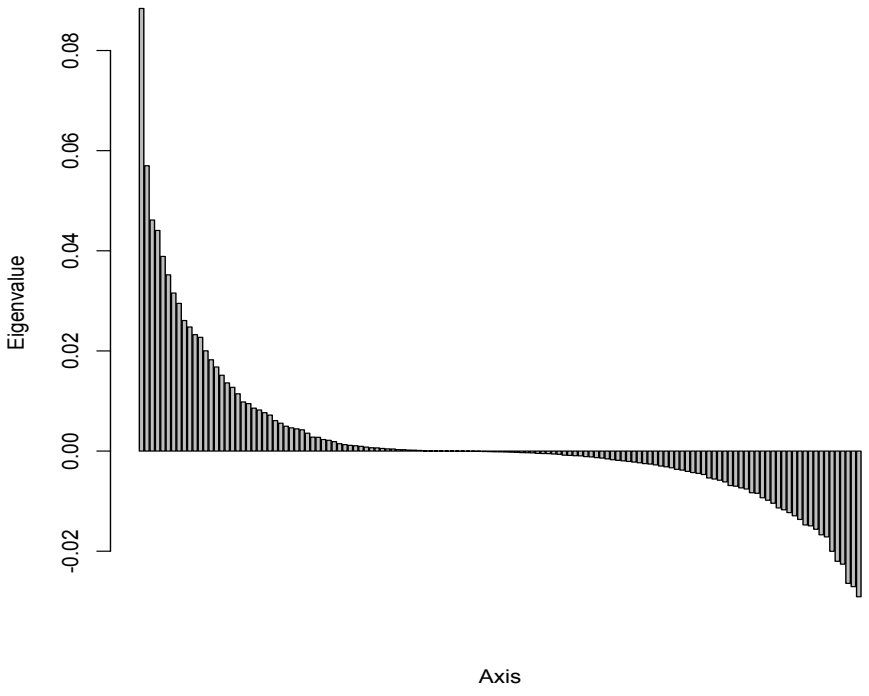


Table 2.1. Locus-specific summary of genetic variation for n = 233 badgers in Wisconsin. Metrics included are the number of alleles per locus (A), observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficients (F_{IS}), and 95% confidence intervals for F_{IS} for each locus. Asterisks indicate significant deviations from HWE at a corrected $\alpha = 0.012$.

Locus	A	H_O	H_E	F_{IS}	F_{IS} Lower CI	F_{IS} Upper CI
Mel111	11	0.68	0.85	0.198*	0.127	0.268
Mel14	10	0.81	0.83	0.022	-0.035	0.083
Mel1	13	0.81	0.87	0.071*	0.012	0.126
Tt-1	13	0.79	0.81	0.024	-0.041	0.084
Tt-2	7	0.59	0.68	0.133*	0.044	0.219
Ma-1	9	0.68	0.68	0.000	-0.079	0.083
Tt-3	15	0.77	0.88	0.126*	0.063	0.187
Tt-4	14	0.83	0.88	0.064*	0.008	0.122
Mvis072	9	0.64	0.81	0.209*	0.136	0.283
Mel101	15	0.81	0.86	0.048	-0.008	0.102
Mel108	13	0.73	0.78	0.064	-0.008	0.123
Mel112	13	0.60	0.81	0.261*	0.184	0.337
Overall	11.83	0.73	0.81	0.103*	0.080	0.122

Table 2.2. Model selection results for spatially lagged regressions on sPCA Axes 1 and 2. Top models (ΔAIC < 2.0; in bold) for Axis 1 all included the Wisconsin River (River) whereas models for Axis 2 incorporated Agriculture (Ag). Ecoregion (Eco) was not significant for either axis following model averaging.

Model	AIC	ΔAIC	w_i
Axis 1			
Eco+River	**169.66**	**0.00**	**0.499**
Ag+Eco+River	**171.46**	**1.80**	**0.203**
River	**171.49**	**1.83**	**0.200**
Ag+River	172.94	3.28	0.097
Ag	199.40	29.74	0.000
Ag+Eco	201.15	31.49	0.000
Eco	203.11	33.45	0.000
Null	204.56	34.90	0.000
Axis 2			
Ag	**216.68**	**0.00**	**0.482**
Ag+River	**218.20**	**1.52**	**0.225**
Ag+Eco	**218.52**	**1.84**	**0.192**
Ag+Eco+River	219.90	3.22	0.096
Eco	227.58	10.90	0.002
Null	229.16	12.48	0.000
River	229.93	13.25	0.000

Table S2.1. Details of the microsatellite loci to be used in this study. Each multiplex set consisted of three loci. Locus amplification conditions used in this study are listed as footnotes.

Locus Name	Multiplex Set[a, b]	Primer Sequence	Primer (pmol)	Size Range
Mel111	1	F: 6-FAM-TGCATACAGCTCCCTGAAAG R: GTGGTAGATGCTGGGATAGTG	8	135-155
Mel14	1	F: HEX-GACACAAGCAAACTTTCTTCC R: CATCTGGAAGTAGGCATAATG	5	169-191
Mel1	1	F: 6-FAM-CTGGGGAAAATGGCTAAACC R: AATGCAGGCTTTGCAATTCC	5	240-276
Mel101	2	F: 6-FAM-ACGGTCCACCAATGATGAAT R: ACAAATGGGAAGGTGTCCT	8	119-149
Me1108	2	F: 6-FAM-GTCTGGAGCCCCATGTTG R: TCTTTGGAATGGAAGTTAATGG	5	330-360
Mel112	2	F: HEX-GATCAAGTCCCACATTGCG R: AAGGTCCATCCATGGTGTTG	10	387-411
Tt-1	3	F: 6-FAM-AACGGCTTCTAACCACTCCA R: CCCCGCTTTTCATTTCTTTA	5	150-176
Tt-2	3	F: HEX-AGCCAAGACACAGAAACAAC R: TTCAAGGATTCAAGGACCAT	5	194-212
Ma-1	3	F: 6-FAM-ATTTTATGTGCCTGGGTCTA R: TTATGCGTCTCTGTTTGTCA	5	190-206
Tt-3	4	F: HEX-GGTGAGACCCTGGAAATAGAAA R: GCTAACCAAACCTACGCAATGAT	8	147-179
Tt-4	4	F: 6-FAM- GGTGAGACCCTGGAAATAGAAA R: GCTAACCAAACCTACGCAATGAT	5	164-196
Mvis072	4	F: HEX-CTGCAAAGCTTAGGAATGGAGA R: CCACTACACTGGAGTTTCAGCA	8	297-339

[a]Thermocycler profile for multiplex sets 1 and 2: Initial denaturation step at 94ºC for 5 min; 21 cycles of 94ºC for 1 min, 65ºC for 1 min (-0.5ºC per cycle), 72ºC for 1 min; 15

cycles of 94ºC for 1 min, 50ºC for 1 min, 72ºC for 1 min; final extension at 72ºC for 2 min; soak at 60ºC for 45 min.

[b]Thermocycler profile for multiplex sets 3 and 4: Initial denaturation step at 94ºC for 5 min; 30 cycles of 94ºC for 30 s, 58ºC for 30 s, 72ºC for 30 s; final extension at 72ºC for 10 min; soak at 60ºC for 45 min.

Literature Cited

Anderson CD, Epperson BK, Fortin MJ, Holderegger R, James PMA, M. S. Rosenberg, K. T. Scribner, and S. Spear 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, **19**, 3565-3575.

Apps CD, Newhouse NJ, Kinley TA (2002) Habitat associations of American badgers in southeastern British Columbia. *Canadian Journal of Zoology*, **80**, 1228-1239.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, **32**, 818-830.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165-1188.

Bivand R, Altman M, Anselin L, *et al*. (2011) R Package 'spdep': spatial dependence: weighting schemes, statistics and models. Available from http://cran.r-project.org/web/packages/spdep/index.html (accessed March 13 2013).

Blanchong JA, Samuel MD, Scribner KT, Weckworth BV, Langenberg JA, Filcek KB (2008) Landscape genetics and spatial distribution of chronic wasting disease. *Biology Letters*, **4**, 130-133.

Bowen BW, Bass AL, Soares L, Toonen RJ (2005). Conservation implications of complex population structure: lessons from the loggerhead turtle (*Caretta caretta*). *Molecular Ecology*, **14**, 2389–2402.

Broquet T, Ray N, Petit E, Fryxell JM, Burel F (2006) Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landscape Ecology*, **21**, 877-889.

Burnham KP, Anderson DR (2002) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Verlag, New York, NY, USA.

Carpenter PJ, Dawson DA, Greig C, Parham A, Cheeseman CL, Burke T (2003) Isolation of 39 polymorphic microsatellite loci and development of a fluorescently labeled marker set for the Eurasian badger (*Meles meles*) (Carnivorea: Mustelidae). *Molecular Ecology*, *Notes* **3**, 610-615.

Carr D, Bowman J, Wilson PJ (2007) Density-dependent dispersal suggests a genetic measure of habitat suitability. *Oikos*, **116**, 629-635.

Cegelski CC, Waits LP, Anderson NJ, Flagstad O, Strobeck C, Kyle CJ (2006) Genetic diversity and population structure of wolverine (*Gulo gulo*) populations at the southern

edge of their current distribution in North America with implications for genetic viability. *Conservation Genetics*, **7**, 197–211.

Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9**, 539.

Croteau EK, Heist EJ, Nielsen CK (2010) Fine-scale population structure and sex-biased dispersal in bobcats (*Lynx rufus*) from southern Illinois. *Canadian Journal of Zoology*, **88**, 536-545.

Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *American Naturalist*, **168**, 486-499.

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology*, **19**, 3592-3602.

Davis CS, Strobeck C (1998) Isolation, variability, and cross-species amplification of polymorphic micro-satellite loci in the family Mustelidae. *Molecular Ecology* **7**:1776-1778.

Domingo-Roura X, Macdonald DW, Roy MS, Marmi J, Terradas J, Woodroffe R, Burke T, Wayne RK (2003) Confirmation of low genetic diversity and multiple breeding females in a social group of Eurasian badgers from microsatellite and field data. *Molecular Ecology*, **12**, 533-539.

Duffy AJ, Landa A, O'Connell M, Stratton C, Wright JM (1998) Four polymorphic microsatellites in wolverine, *Gulo gulo*. *Animal Genetics*, **29**, 63.

Duquette JF, Gehrt SD (2014) Badger (*Taxidea taxus*) resource selection and spatial ecology in intensive agricultural landscapes. *The American Midland Naturalist*, **171**, 116-127.

Epperson BK, McRae BH, Scribner K, Cushman SA, Rosenberg MS, Fortín M, James PMA, Murphy M, Manel S, Legendre P, Dale MRT (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology*, **19**, 3549-3564.

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

Fleming MA, Ostrander EA, Cook JA (2002) Microsatellite markers for American mink (*Mustela vison*) and ermine (*Mustela erminea*). *Molecular Ecology*, **8**, 1352-1355.

Fortín MJ, Legendre P (2010) Methodological advances–inference of population structure: comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology*

Resources, **10**, 831–844.

Frantz AC, Pope LC, Carpenter PJ, Roper TJ, Wilson GJ, Delahay RJ, Burke T (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology*, **12**, 1649-1661.

Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493-505.

Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L, Barnes C, Herold N, Wickham J (2011) Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, **77**, 858-864.

Garroway CJ, Bowman J, Wilson PJ (2011) Using a genetic network to parameterize a landscape resistance surface for fishers, *Martes pennanti*. *Molecular Ecology*, **20**, 3978-3988.

Goldberg CS, Waits LP (2010) Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. *Molecular Ecology*, **19**, 3650-3663.

Graves TA, Beier P, Royle A (2013) Current approaches using genetic distances produce poor estimates of landscape resistance to inter-individual dispersal. *Molecular Ecology*, **22**, 3888-3903.

Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods in Ecology and Evolution*, **4**, 336-344.

Hanski I (1999) Metapopulation ecology. Oxford University Press, Oxford, UK.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Hole FD (1976) Soils of Wisconsin. University of Wisconsin Press, Madison, Wisconsin, USA.

Hoodicoff CS (2003) Ecology of the badger (*Taxidea taxus jeffersoni*) in the Thompson region of British Columbia: implications for conservation. *M.Sc. Thesis*. University of Victoria, Victoria, Canada.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801-1806.

Johansson M, Primmer CR, Merilla J (2007) Does habitat fragmentation reduce fitness and adaptability? A case study of the common frog (*Rana temporaria*). *Molecular Ecology*, **16**, 2693-2700.

Jombart T, Devillard S, Dufour A, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92-103.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405.

Kamler JF, Gray MM, Macdonald DW (2013) Genetic structure, spatial organization, and dispersal in two populations of bat-eared foxes. *Ecology and Evolution*, **3**, 2892-2902.

Keenan K (2013) diveRsity: Genetic diversity partition statistics and informative locus selection using Fst, Gst, Dest (Jost Chao) G'st and In. Version 1.6.1. http://diversityinlife.weebly.com/ (accessed 20 April 2014).

Kierepka EM, Latch EK, Swanson BJ (2012) Influence of sampling scheme on the inference of sex-biased gene flow in the American badger (*Taxidea taxus*). *Canadian Journal of Zoology*, **90**, 1231-1242.

Kitchell J (2008) Wisconsin wildlife harvest summary 1930-2007. Wisconsin Department of Natural Resouces, Madison, Wisconsin, USA.

Landguth EL, Cushman SA (2009) CDPOP: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, **10**, 156-161.

Landguth EL, Cushman SA, Schwartz MK, McKelvey KS, Murphy M, Luikart G (2010) Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, **19**, 4179-4191.

Latch EK, Boarman WI, Walde A, Fleischer RC (2011) Fine-scale analysis reveals cryptic landscape genetic structure in desert tortoises. *PloS One*, **6**, e27794.

Legendre P, Legendre L (1998) Numerical Ecology. Elsevier Science, Amsterdam, Netherlands.

Lindzey FG (1978) Movement patterns of badgers in northwestern Utah. *Journal of Wildlife Management*, **42**, 418-422.

Messick JP, Hornocker MG (1981) Ecology of the badger in southwestern Idaho.*Wildlife Monographs*, **76**, 1-53.

Ministry of Environment Ecosystems Branch for the Resources Information Standards Committeee (2007) Inventory methods for medium-sized territorial carnivores: badger.

Standards for Components of British Columbia's Biodiversity No. 25a. ISBN 978-0-7726-5736-7.

Navidi W, Arnheim N, Waterman MS (1992) A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics*, **50**, 347-359.

Oksanen J, Kindt R, Legendre P, O'Hara RB (2008) vegan: community ecology package. Version 1.7–81. http://cran.r-project.org/ (accessed on 15 February 2013).

Omernik JM (1987) Ecoregions of the conterminous United States. Map (scale 1:7,500,000). *Annals of the Association of American Geographers*, **77**, 118-125.

Oyler-McCance SJ, Fedy BC, Landguth EL (2013) Sample design effects in landscape genetics. *Conservation Genetics*, **14**, 275-285.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

QIAGEN (2006) User developed protocol: purification of total DNA from nails, hair, or feathers using the DNeasy® Blood and Tissue Kit. QIAGEN-DYO4-Aug06. (accessed on 5 August 2010).

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/ (accessed on 2 January 2013).

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

Reding DM, Bronikowski AM, Johnson WE, Clark WR (2012) Pleistocene and ecological effects on continental-scale genetic differentiation in the bobcat (*Lynx rufus*). *Molecular Ecology*, **21**, 3078-3093.

Robinson SJ, Samuel MD, Lopez DL, Shelton P (2012) The walk is never random: subtle landscape effects shape gene flow in a continuous white-tailed deer population in the midwestern United States. *Molecular Ecology*, **21**, 4190-4205.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62

Row JR, Gomez C, Koen EL, Bowman J, Murray DL, Wilson PJ (2012) Dispersal promotes high gene flow among Canada lynx populations across mainland North America. *Conservation Genetics*, **13**, 1259-1268.

Row JR, Wilson PJ, Gomez C, Koen EL, Bowman J, Thornton D, Murray DL (2014) The subtle role of climate change on population genetic structure in Canadian lynx. *Global Change Biology*, **20**, 2076-2086.

Schwartz MK, McKelvey KS (2010) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441-452.

Schwartz MK, Copeland JP, Anderson NJ, Squires JR, Inman RM, McKelvey KS, Pilgrim KL, Waits LP, Cushman SA (2009) Wolverine gene flow across a narrow climatic niche. *Ecology*, **90**, 3222-3232.

Segelbacher G, Cushman SA, Epperson BK, Fortín M, Francois O, Hardy OJ, Holderegger R, Taberlet P, Waits LP, Manel S (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, **11**, 375-385.

Shafer ABA, Northrup JM, White KS, Boyce MS, Cote SD, Coltman DW (2012) Habitat selection predicts genetic relatedness in an alpine ungulate. *Ecology*, **93**, 1317-1329.

Short Bull RA, Cushman SA, Mace R, Chilton T, Kendall KC, Landguth EL, Schwartz MK, McKelvey K, Allendorf FW, Luikart G (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology*, **20**, 1092-1107.

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

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**, 3189-3194.

Tammeleht E, Remm J, Korsten M, Davison J, Tumanov I, Saveljev A, Männil P, Kojola I, Saarma U (2010) Genetic structure in large, continuous mammal populations: the example of brown bears in northwestern Eurasia. *Molecular Ecology*, **19**, 5359-5370.

Walker CW, Vilà C, Landa A, Linden M, Ellegren H (2001) Genetic variation and population structure in Scandinavian wolverine (*Gulo gulo*) populations. *Molecular Ecology*, **10**, 53-63.

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, **67**, 3403-3411.

Warner RE, Ver Steeg B (1995) Illinois badger studies. Division of Wildlife Resources. Department of Natural Resources and Environmental Sciences, Springfield, Illinois, USA.

Wisconsin Department of Natural Resources 2008. Wisconsin's Wildlife Action Plan (2005-2015). Madison, Wisconsin, USA.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114-138.

Wydeven AP, Wiedenhoeft JE, Dhuey B (1998) Status of badgers in Wisconsin 1987-1998. Wisconsin Department of Natural Resources, Madison, Wisconsin, USA.

Zeller KA, McGarigal K, Whiteley AR (2012) Estimating landscape resistance to movement: a review. *Landscape Ecology*, **27**, 777-797.

CHAPTER 3: RANGE-WIDE GENE FLOW IN A WIDESPREAD, SPECIALIST CARNIVORE, THE AMERICAN BADGER (*TAXIDEA TAXUS*)

Introduction

Contemporary species distributions are the direct result of the geological, climatic, and ecological conditions that influenced evolution across landscapes. In North America, evidence suggests that climatic oscillations during the Pleistocene represent a major evolutionary force in North America, particularly for temperate species. Taxa in northern latitudes were displaced by glaciations and restricted to one or more refugia, which often results in genetic variation that reflects population fluctuations and isolation within refugia during the Last Glacial Maximum (LGM; Hewitt 2000). Comparative studies in biogeographic hotspots such as the Pacific Northwest (Shafer *et al*. 2010) have identified common ecological factors that impact how co-distributed species responded to Pleistocene glaciation. Among these ecological factors, habitat specialization appears to play a key role in location and number of refugia occupied as well as patterns of recolonization after glaciation, a finding that is consistent with Pleistocene fossil records (e.g., Graham *et al*. 1996).

Pleistocene glaciations displaced many temperate species, and resultant distributional ranges and genetic structure in taxa with strong habitat preferences often closely mirror the changes that occurred within their preferred habitat type. For example, many forest-associated species exhibit eastern and western clades within North America because ice sheets, topographic features like the Rocky Mountains, and aridification

within the Great Plains separated areas of suitable forest (e.g., Wooding & Ward 1997, Weir & Schluter 2004, Runck & Cook 2005). While the Great Plains acted as a barrier to forest species, many grassland specialists were restricted to central North America and then expanded when habitats became more suitable during the Holocene (e.g., Johnson *et al*. 2003; Oyler-McCance *et al*. 2005; Wisely *et al*. 2008). According to fossil records during the climatic oscillations of the Quaternary, at the LGM, assemblages of species within habitats were highly predictable because they were largely defined by similar habitat preferences (Graham *et al*. 1996). For strict specialists, the distribution of preferred habitats likely continued to be the dominant driver of dynamics until contemporary time scales because strict specialists cannot persist outside their suitable habitat. Consequently, many strict specialists exhibit genetic structure that reflects both historical and contemporary changes in their preferred habitat (Curtois *et al*. 2003; Wisely *et al*. 2008; Shafer *et al*. 2011; Trumbo *et al*. 2013).

Outside strict specialists, previously defined assemblages of organisms during the LGM became highly variable following glacial recession (Graham *et al*. 1996), presumably because other species-specific traits contributed to post-glacial expansions into newly available habitats. For example, high dispersal capabilities allow individuals to traverse potential barriers and colonize new habitats quickly, so many highly mobile species exhibit patterns associated with rapid expansion following glacial recession (Vila *et al*. 1999; Latch *et al*. 2009; Reding *et al*. 2012). High dispersal capabilities also can facilitate varying levels of homogenization of previously isolated refugial populations (Latch *et al*. 2009; Smith *et al*. 2011) and even limited genetic structure in specialists within heterogeneous landscapes (e.g., Garroway *et al*. 2011; Centeno-Cuadros *et al*.

2011). However, high dispersal capabilities do not always prevent genetic differentiation, particularly when accelerated genetic drift (i.e., low effective population size; Schwartz *et al*. 2009; Weckworth *et al*. 2014) or limited realized dispersal (e.g., natal habitat biased dispersal; Sacks *et al*. 2005, isolation-by-ecology; Musiani *et al*. 2007; Louis *et al*. 2014) resist the homogenizing effects of gene flow. Therefore, even if species retain signatures of past isolation within glacial refugia that are common among species with similar habitat preferences (e.g., Aubry *et al*. 2009; Barton *et al*. 2012; Reding *et al*. 2012; Latch *et al*. 2014), predicting overall genetic structure is difficult due to the species-specific responses that have occurred according to changes in habitat following the LGM into contemporary timescales.

One group that exemplifies the difficulty in understanding how habitat specialization influences gene flow are mammalian carnivores because they typically have large geographic ranges and high dispersal capabilities combined with low population densities and substantial variation in morphology and behavior throughout their range. Large geographic ranges and high dispersal capabilities predict extensive gene flow whereas low densities and morphological/behavioral variation suggest high rates of genetic drift and differential selection, respectively. Many carnivores are subject to intense management, so understanding genetic structure in carnivores has received considerable attention in the literature, particularly in forest or arctic-associated species (e.g., Schwartz *et al*. 2009, Short Bull *et al*. 2011; Koen *et al*. 2012; Reding *et al*. 2012; Row *et al*. 2012). From these studies, a wide array of responses to habitat heterogeneity have been recorded including differentiated ecotypes (Musiani *et al*. 2007), landscape variables driving genetic differentiation (Schwartz *et al*. 2009; Short Bull *et al*. 2011;

Reding *et al*. 2012), and continent-wide patterns of isolation-by-distance (Tammeleht *et al*. 2010; Row *et al*. 2012). Unlike forest and arctic-associated carnivores, grassland species have received considerably less attention in genetic studies (but see Wisely *et al*. 2008), yet they can provide insight into how changes during and following the LGM such as forest encroachment and massive conversion to agriculture have contributed to genetic variation.

This study focused on the American badger (*Taxidea taxus*), a semifossorial, grassland-associated mesocarnivore that has a large geographic range encompassing most of central and western North America where suitable grassland habitats occur. Along with strong preferences for grassland habitats, their semifossorial lifestyle makes badgers dietary specialists as the vast majority of their diet is composed of available burrowing mammals like prairie dogs and ground squirrels (Messick and Hornocker 1981; Goodrich and Buskirk 1998; Sovada *et al*. 1999; Azevedo *et al*. 2006). Evidence for differentiation in badgers across their range is reflected in subspecific designations as morphological and behavioral differences define four subspecies (*T. t. berlandieri*, *T. t. jacksoni*, *T. t. jeffersonii*, and *T. t. taxus*; Long 1972; Figure 3.1). Each subspecies roughly corresponds to grassland habitat types (prairie, scrub-steppe, and oak savannas) or large physical barriers (Rocky Mountains, Grand Canyon; Long 1972). Taken together, specialist traits and apparent differences according to habitat (i.e., subspecific designation) predict limited gene flow between habitat types.

Although American badgers possess traits that suggest habitat-based differentiation, their large dispersal capabilities and geographic range are more typical of a generalist. Their large geographic range greatly exceeds other grassland specialists and

includes a number of hypothesized glacial refugia for mammals (Great Plains; Wisely *et al*. 2008, Pacific Northwest; e.g., Latch *et al*. 2009, Shafer *et al*. 2010; Great Basin and California; Howard & Swenson 2005). Pleistocene badger fossils have been found throughout North America (Long 1972), suggesting that badgers may not have been as limited in distribution as other highly specialized species during the LGM. Badgers are also highly mobile as demonstrated by their seasonally large home ranges and dispersal distances (125 km+; Messick & Hornocker 1981). Therefore, high mobility and corresponding gene flow in badgers could theoretically override any isolation created by habitat preferences or Pleistocene glaciation.

In this study, we provide a continent-wide assessment of genetic structure within the American badger. Our primary goal was to investigate if specialization on grassland habitats impacted dynamics during Quaternary glaciation cycles and contemporary gene flow. If specialization on grasslands drives gene flow patterns in badgers, badgers would be expected to have experienced a restricted distribution in the Great Plains (sensu Johnson *et al*. 2003; Oyler-McCance *et al*. 2005; Wisely *et al*. 2008) with subsequent rapid recolonization following glacial recession. Also, subspecies designations should be correlated with genetic differentiation if specialization impacts contemporary gene flow. Conversely, if high mobility overrides any effect of specialization, badgers are expected to exhibit limited genetic differentiation except when confronted with large topographic barriers. To test these predictions, we sampled 917 badgers from across North America representing all four subspecies to evaluate the effect of topographic barriers, geographic distance and ecological variables on genetic differentiation. This broad scale genetic survey can aid in understanding how ecological traits such as specialization and mobility

interact to produce gene flow patterns in grassland species. Additionally, specialists like badgers are vulnerable to fragmentation (Crooks 2002) and grasslands have undergone intense conversion to agriculture, so observed genetic patterns may also provide insight into how fragmentation has affected connectivity in badger populations.

Methods

Sample Collection

We collected 917 American badgers from 2001-2013, representing all four described subspecies and the majority of the North American range (Figure 3.1). Samples included tissue (n = 520) and skin (n = 397) collected from road-killed and fur-trapped animals (Table S3.1). All samples were georeferenced, either as latitude/longitude coordinates recorded at the time of sample collection or estimated based on written descriptions of the sampling locations. Exact locations for location descriptions (nearest city, state/province, Public Land Survey System [PLSS]) were calculated in ArcGIS 10.1 using the "create random points" function either within a political boundary (PLSS, state, county, or province) or within 20km circular buffers around a city center (buffer radius = 5, 10, 20, or 50 km yielded concordant results).

Laboratory Methods

DNA was extracted from all samples and all sample types using the Qiagen DNEasy Blood and Tissue Kit. A 600 bp segment of the mitochondrial displacement loop (d-loop) was amplified using newly designed primers (Tax-dloopF 5'-

CGGGGTCTTGTGACTCTTCT-3' and Tax-dloopR 5'-CAGCACCCAAAGCTGATATTC-3'). PCR amplifications occurred in 10 uL reactions with 10 ng of genomic DNA, 5 pmol of each primer, 0.2 mM of each dNTP, 0.75 U of PerfectTaq DNA polymerase, and 1 X PerfectTaq PCR Buffer. Thermocycler conditions for mtDNA included a 2-min initial denaturation step at 94°C, 30 cycles of 30 second denaturation (94°C), annealing (60°C), and extension (72°C) steps, and final extension at 72°C for 5 min. Each 10 uL sequencing reaction contained 50 ng of template, 0.3 pmol of forward primer, 0.5 uL of ABI Big Dye in 2 X reaction buffer. Thermocycler conditions followed the manufacturer's recommendations, except 60 cycles were used to improve signal strength. Cleaning followed a modified low sodium precipitation method (Latch & Rhodes 2005), and then cleaned products were sequenced on an ABI3730 DNA Analyzer at the University of Wisconsin Biotechnology Center. All sequences were analyzed using the program GENEIOUS PRO v. 5.2.

For quality control, we re-extracted and re-sequenced all samples yielding unique haplotypes (n = 54), all samples containing a 25 bp deletion (n = 56), and a randomly selected subset of remaining badgers (n = 103). We observed four sequencing errors (4/213 = 1.88% error rate); three in individuals with unique haplotypes and one in an individual with a common haplotype. A representative sequence for each unique haplotype was deposited in Genbank. Our final mtDNA dataset contained 785 badgers with mtDNA sequences, which excluded all individuals with the 25 bp deletion region (n = 14 haplotypes with deletion).

We amplified all samples at 12 microsatellite loci developed in American badger (*Taxidea taxus*; Tt-1, 2, 3, and 4; Davis & Strobeck 1998), American mink (*Neovison*

vison; Mvis072; Fleming *et al*. 2002), American marten (*Martes americana*; Ma-1; Davis & Strobeck 1998), and European badger (*Meles meles*; Mel112, Mel101, Mel111, Mel108, Mel14, and Mel1: Carpenter *et al*. 2003; Domingo-Roura *et al*. 2003). PCRs were conducted in four multiplex reactions containing 3 primers each (conditions described in Chapter 2). Each 10 uL reaction contained 10 ng of genomic DNA, 3-8 pmol each primer, 0.2 mM of each dNTP, 0.75 U of PerfectTaq DNA polymerase, and 1 X PerfectTaq PCR Buffer. Amplified products were genotyped on an ABI3730 DNA Analyzer at the University of Wisconsin Biotechnology Center, and alleles were sized using the program GeneMarker (SoftGenetics LLC, State College, PA, USA).

For quality control, we re-genotyped all individuals that had rare alleles (i.e., occurred in fewer than 2% of samples; n = 17) and those in mtDNA quality control (n = 213) to calculate error rates and confirm rare alleles. These repeated PCRs and amplifications confirmed all rare alleles and detected one mismatch due to probable allelic dropout (error rate = 1/2760; 0.0036%). All individuals (n = 917) were successfully genotyped at 10 or more loci (11/11,004 total genotypes missing; 0.10% missing).

Historical Gene Flow

To investigate historical patterns in gene flow, we generated phylogenetic trees using three algorithms (Maximum Parsimony; MP, Maximum Likelihood; ML, and Bayesian) implemented in GENEIOUS. The GENEIOUS plug-in for PAUP (Swofford 2003) and MrBayes (Ronquist *et al*. 2012) performed calculations identical to their parent programs with pre-defined parameters. We defined gaps as a fifth state and used the 50%

majority rule consensus trees for all three algorithms. jMODELTEST 2 (Darriba *et al*. 2012) suggested that the GTR+G+I mutation model (Tavaré 1986) with a gamma shape of 0.49 best fit our data. A *Meles meles* d-loop sequences (AM711900.1) served as the outgroup for all analyses. Support nodes within MP and ML (bootstraps) as well as Bayesian trees (posterior probabilities) were calculated after 1000 permutations.

Traditional tree-building techniques can yield unresolved intraspecific trees (Posada & Crandall 2001), so a haplotype network was constructed to better visualize relationships among haplotypes. The network was assembled in the program TCS (Clement *et al*. 2000) based on maximum parsimony. We used a 95% connection limit with gaps as a 5th state. The resultant haplotype network connected all haplotypes into a single network.

To complement tree-building and network analyses, we also tested for differentiation in mtDNA using a spatial analysis of molecular variance (SAMOVA; Dupanloup *et al*. 2002). SAMOVA identifies the number of groups of populations (*K*) that maximizes ϕ_{CT} or total genetic variance explained by differentiation between groups. SAMOVA requires individuals be grouped into populations, so we grouped badgers into a priori populations (n = 19 or 23; Table S3.1). For half of the Canadian samples, we had only location of fur deposit, not the specific trapping location, so SAMOVAs were run with either all Canadian badgers pooled as 'Canada' or only including individuals that had specific trapping locations (Saskatchewan = 27, Manitoba = 5, Alberta = 5). To generate geographic coordinates for each population, we drew 100% minimum convex polygons around geo-referenced individuals within each state or Canada via the Minimum Bounding Geometry function in ArcGIS. Latitude and longitude points were

calculated for the centroid of each polygon, and used as the geographic coordinates for each population within the SAMOVA. We tested $K = 2$ through 10 using the program SAMOVA v. 1.0 (Dupanloup *et al*. 2002).

We used ARLEQUIN 3.5 (Excoffier & Lischer 2010) to calculate basic diversity metrics for the total dataset, four subspecies, and SAMOVA groups. Diversity metrics for each group included haplotype diversity (h), nucleotide diversity (π), and number of pair-wise differences. Differentiation (ϕ_{ST}) between SAMOVA groups and subspecies were also quantified in ARLEQUIN.

To test for evidence of postglacial demographic expansion, we used ARLEQUIN to calculate three complementary statistics. Fu's F_S (Fu 1997) and Tajima's D (Tajima 1989) were calculated for the total dataset, subspecies, and SAMOVA groups. Significantly negative values of F_S and D indicate rapid demographic expansion. We further evaluated evidence for postglacial expansion by constructing mismatch distributions of nucleotide differences for all individuals, subspecies, and SAMOVA groups. Under constant population size, mismatch distributions exhibit a ragged pattern whereas recently expanded populations have a unimodal distribution (Slatkin & Hudson 1991; Rogers & Harpending 1992). To assess fit of the observed nucleotide differences to expected distributions under sudden demographic or spatial expansion, we calculated two statistics: raggedness index (Rogers & Harpending 1992) and sum of squared differences (SSD; Schneider & Excoffier 1999) in ARLEQUIN.

Another method to examine recolonization dynamics following glacial recession is through simulations that attempt to reconstruct demographic histories (e.g., Heled & Drummond 2008). However, the requirement for specific priors, particularly lineage

specific mutation rates or fossil calibration points, would have yielded considerable uncertainty in badgers. When we calculated mutation rates for badgers, a highly divergent, monotypic lineage in Mustelidae (approximately 20 mya diverged from common ancestor of remaining mustelids; Koepfli *et al*. 2007), mutation rates were very low (e.g., 3.835×10^{-9} bp/generation using *Meles* clade) compared to other estimates of d-loop in mammals (e.g., Koblmüller *et al*. 2012). Regions of d-loop typically have a higher mutation rate compared to the rest of the mitogenome and tend to have species-specific mutation rates (Pesole *et al*. 1999). Therefore, we hypothesize that mutations within d-loop obscured the distinctiveness of the American badgers through homoplasy, similar to d-loop in fisher (*Martes pennanti*; Knaus *et al*. 2011). Inaccurate mutation rates has been recorded when using ancient divergences to calibrate nodes for demographic reconstruction (e.g., Ho *et al*. 2005; Navascués & Emerson 2009), and using different mutation rates or fossils for calibration can give radically divergent conclusions (e.g., Ho *et al*. 2005; Shapiro & Ho 2014). Based on the likely inaccurate mutation rate for d-loop and little differentiation among haplotypes, we chose not to estimate divergence times and demographic parameters in this dataset.

Contemporary Gene Flow

Two Bayesian programs, non-spatial STRUCTURE 2.2.3 (Pritchard *et al*. 2000) and spatial BAPS 5 (Corrander *et al*. 2008), were run to examine contemporary population structure of badgers in North America. In our non-spatial approach, we performed clustering in STRUCTURE for all 917 individuals for 5 iterations of each value of *K* from *K* = 1 – 15. Each run consisted of 100,000 replicates of the MCMC after a burn-in of

100,000 replicates. We used both likelihoods and Evanno *et al*. (2005)'s ΔK method to determine the most likely *K* since likelihoods plateaued and exhibited higher variances after the optimum *K* is reached (Pritchard *et al*. 2000; Pritchard & Wen 2003). Once the optimum *K* was identified, longer runs (1,000,000 MCMC burn-in, 1,000,000 permutations) were performed at *K*-1, *K*, and *K*+1 to verify results from the shorter runs. To assign individuals to inferred clusters, we performed a final set of 10 runs (1,000,000 burn-in, 1,000,000 stored replicates) at the optimal *K*. Each individual was assigned to the cluster in which it had the highest average probability of membership, based on averaged *q*-values (i.e., proportion of genome that belonged to each cluster) calculated in CLUMPP (Jakobsson & Rosenberg 2007). The entire clustering process was repeated iteratively for each of the inferred clusters, in order to identify any additional substructure within the main genetic clusters.

In our spatially explicit approach, we employed Bayesian clustering in BAPS 5 using the "spatial clustering of individuals" model. We performed five iterations for each value of *K* for *K* = 1-15. The maximum likelihood and highest posterior probability were used to determine the optimum number of genetic clusters in the sample, and each individual was assigned to a cluster. We then performed an admixture analysis based on the results of the mixture clustering, using 100 iterations, 20 reference individuals per population, and 5 iterations per reference individual within clusters.

The Bayesian algorithms yielded the same *K* following iterative runs in STRUCTURE, but the locations of genetic discontinuities between clusters were more defined in BAPS. Therefore, we utilized results from BAPS to examine genetic diversity and differentiation between putative clusters. Deviations from Hardy-Weinberg and

Linkage Equilibrium were quantified in GENEPOP (Raymond & Rousset 1995) for each cluster and the total dataset following corrections for multiple tests (false-discovery rate; Benjamini & Yekutieli 2001). We used the R package *diveRsity* (Keenan 2013) to calculate both genetic differentiation between clusters (F_{ST}) and genetic diversity metrics (heterozygosity, allelic richness, and F_{IS}).

A number of processes likely influence genetic variation in badgers across North America. One such process is isolation-by-distance (IBD; Wright 1941), a phenomenon where genetic differentiation is correlated with geographic distance. To test for IBD, we utilized simple Mantel tests that quantify the correlation (i.e. Pearson's *r*) between genetic distances and Euclidean distances. Pair-wise genetic distances (Rousset's *a*; Rousset 2000) were calculated in SPAGEDI (Hardy & Vekemans 2002) and Euclidean distances were calculated in ArcMap. Simple Mantel tests were performed in the R package *vegan* (function "mantel"; Oksanen *et al*. 2008) for the entire dataset and for each group identified in the Bayesian clustering analyses, and statistical significance was assessed after 1000 permutations.

In addition to IBD, we tested how three topographic barriers (Rocky Mountains, Mississippi River, and Lake Michigan) and subspecific designations contribute to observed patterns in genetic variation via three tests. Partial Mantel tests, principal components analysis (PCA), and spatial principal components analysis (sPCA) provide complementary methods to test for barrier effects. The first test, partial Mantel tests, control for the effects of geographic distances to evaluate correlations between each topographic barrier or subspecific designation and pair-wise genetic distances. For subspecies, predictor variables were coded as dummy variables (0 or 1) based on whether

individuals were of the same (0) or different (1) subspecies. We only tested differences between adjacent subspecies, resulting in four partial Mantel tests (*T. t. jacksoni/taxus, T. t. taxus/jeffersonii, T. t. taxus/berlandieri, T. t. jeffersonii/berlandieri*; Figure 3.1 for subspecies boundaries). For each partial Mantel test, only individuals within 300 km of the putative barrier or subspecies boundary were included to control for any confounding effects of geography or other barriers. Statistical significance was evaluated via Pearson's correlation coefficients (*r*) after 1000 permutations in the R package *vegan* (function partial.mantel; Oksanen *et al*. 2008).

To complement the partial Mantel tests, we used PCAs to visualize the potential roles of all topographic barriers (Mississippi River, Rocky Mountains, and Lake Michigan), water barriers only (Mississippi River and Lake Michigan), and subspecific designations on gene flow. We summarized patterns in genetic diversity into linear axes called principal components in the dudi.pca function within the R package *adegenet* (Jombart *et al*. 2008). We accounted for missing data by filling in empty genotypes with mean allele frequencies using the "scaleGen" function. For each grouping (topographic barriers, water barriers, or subspecies), 95% inertia ellipses drawn around each group permit visual inspection of the differentiation among groups. In total, we drew 95% ellipses around individuals according to three groupings: topographic barriers (Mississippi River, Rocky Mountains, and Lake Michigan), water barriers only (Mississippi River and Lake Michigan), and subspecific designations (*T. t. berlandieri*, *T.t. jacksoni*, *T. t. jeffersonii*, and *T. t. taxus*).

sPCA provides principal components scores that summarize both the non-spatial genetic variation and the spatial autocorrelation structure among individual genotypes

(Jombart *et al*. 2008). In this method, highly positive eigenvalues reflect axes with large variances and positive spatial autocorrelation (i.e., global patterns; Jombart *et al*. 2008). We performed a sPCA in *adegenet* using a distance-based connection network to ensure all individuals were connected by at least one segment. Badgers were unlikely to be completely genetically isolated, so this network is likely most realistic. To assess which axes are important for explaining genetic variation, we examined two outputs generated from the sPCA. Axes with the highest eigenvalues are considered the most important because they explain the most variance, so eigenvalues were plotted to examine differences among eigenvalues for each axis. Another visualization that aids in interpretation of sPCA results is a screeplot, a graph of spatial autocorrelation vs. variance explained for each axis. Axes that are well differentiated from other axes explain the most variation in genetic structure. Spatially lagged scores from axes that had the highest eigenvalues and were separated from most others in the screeplot were plotted in ArcMap to examine any geographic structuring within the sPCA. We conducted a permutation procedure with 1000 randomizations to test for a significant global ("global.rtest") pattern in the data.

Topographic barriers are likely to be important in shaping badger gene flow, but they may occur in tandem with other landscape or climate variables as observed in other carnivores (e.g., Garroway *et al*. 2011; Reding *et al*. 2012). Therefore, we constructed a full model of how landscape heterogeneity impacts gene flow by including landscape (soil, land cover, and precipitation) variables in addition to the three topographic barriers, subspecies, and geographic coordinates. The first set of landscape variables included soil parameters likely to be important for semifossorial badgers: soil texture (% clay, sand, or

silt) and soil order (5 total within sampled area: United States Department of Agriculture 1999) for each sample location. We also derived land cover (NCLD2006; Fry *et al*. 2011) and precipitation (average annual precipitation from WorldClim v. 1.4; Hijmans *et al*. 2005) data for each sample location. Subspecific designations roughly correspond to major habitat breaks defined by level I ecoregions, regions of similar biotic and abiotic factors within North America (Obernick 1987). Therefore, subspecific designations were included as predictor variables (dummy variables 1-4) as well as the three topographic barriers (Rocky Mountains, Mississippi River, and Lake Michigan) and geographic coordinates (latitude and longitude). In total, ten landscape variables (soil texture, soil order, land cover, average annual precipitation, subspecies, Rocky Mountains, Mississippi River, Lake Michigan, latitude, and longitude) served as predictors within distance-based redundancy analysis (dbRDA) models.

We used dbRDA, a multivariate technique analogous to linear regression, to evaluate the relationship between landscape predictor variables and pair-wise genetic distances (Rousset's *a*). An initial marginal test analyzed all landscape variables in DISTLM v. 5 (Anderson 2004) while subsequent partial dbRDAs calculated the impact of every variable individually. Due to multiple tests, alpha was corrected (alpha = 0.015) for the partial dbRDAs when assessing significance. The program DISTLM FORWARD 3 (Anderson 2003) eliminated non-significant variables via forward selection to produce the best combination of predictors that explained genetic differentiation. By utilizing both partial tests that isolate each variable and forward selection that adds variables sequentially, we were able to control any effects of multicollinearity within our predictor variables.

Results

Historical Gene Flow

In total, 95 haplotypes were identified among 785 individual badgers (54 polymorphic sites, 53 substitutions). The mean number of pairwise differences among sequences was 3.660 (SD = 1.856) and the overall nucleotide diversity was 0.0062 (SD = 0.0035). No *a priori* groupings (four subspecies and two SAMOVA groups) exhibited significant ϕ_{ST} values (all $\phi_{ST} = 0.000$, $p = 0.999$), suggesting limited differentiation in mtDNA in the total sample.

The haplotype network showed some spatial structuring of mtDNA haplotypes. The overall pattern was dominated by at least two common haplotypes surrounded by similar haplotypes, producing star-like patterns (Figure 3.2). The most common haplotype (Haplotype 1, n = 257, 32.7%) occurred throughout North America whereas the second most common (Haplotype 3; n = 149, 19.0%) was primarily found in northern latitudes west of Wisconsin (i.e., North Dakota, Montana, and Canada). The last two common haplotypes (Haplotype 2, n = 96, 12.2%; and Haplotype 4, n = 30, 3.8%) were a single mutation step away from Haplotypes 1 and 3 respectively. Haplotype 2 was largely restricted to Michigan and Ohio (76/96 occurrences) whereas Haplotype 4 was primarily found west of the Rocky Mountains (28/30 occurrences; Figure S3.1). Both Haplotypes 1 and 2 were shared between all four subspecies while Haplotype 3 only occurred in *T. t. jacksoni* and *T. t. taxus* (Figure 3.2a). Haplotype 4 was largely concentrated within *T. t. berlandieri* and *T. t. jeffersonii*. All other haplotypes had a frequency of 10 individuals (1.3%) or less. Despite Haplotypes 1 and 2 exhibiting star-

like patterns often consistent with recent bottlenecks followed by rapid expansion, little differentiation was observed between mitochondrial haplotypes. All tree-building approaches produced phylograms that had a comb-like topology where only the outgroup (*Meles meles*) received high support (bootstraps > 80, posterior probabilities > 0.80), further supporting limited differentiation among haplotypes.

The SAMOVA found the optimum split in haplotypes occurred at $K = 2$ regardless if Canadian provinces were pooled or separate, which corresponded to the Lower Peninsula of Michigan (LP), Indiana, and Ohio versus all other individuals (Table 3.1). Additional evidence for $K = 2$ being the best grouping to explain variance in our mtDNA was that the fixation index (ϕ_{CS}) peaked at $K = 2$, and in simulated datasets, ϕ_{CS} reached its maximum at the appropriate K (Dupanloup *et al*. 2002). The cause of differentiation was driven by the high frequency of Haplotype 3 within the LP, Indiana and Ohio (Figure 3.2b).

All three tests for population expansion supported rapid population expansion in the entire dataset, each subspecies, and western SAMOVA groups. Fu's F_S ($F_S = -24.197$, $p < 0.001$) and Tajima's D ($D = -1.395$, $p = 0.043$) were significantly negative in the total dataset whereas only Fu's Fs remained significant in subspecies and SAMOVA groups (all $p < 0.002$; Table 3.1). Both raggedness indices (all $r < 0.0928$, $p > 0.052$) and SSD (all SSD < 0.313, $p > 0.19$) suggested that the observed mismatch distributions followed predicted distributions under rapid or spatial expansion for all except the eastern SAMOVA group (Figure 3.3). Raggedness and SSD for the eastern SAMOVA group rejected both rapid and spatial expansion ($r = 0.209$, $p = 0.002$; SSD = 0.028, $p = 0.008$). However, the number of haplotypes within the Eastern SAMOVA group was very small

(n = 5), suggesting that the statistically significant results in raggedness and SSD could be due to low sample size. Metrics based on mismatch distributions tend to be less powerful than Fu's F_S (Ramos-Onsins & Rozas 2002), so the significant Fu's F_S indicates that the significance in raggedness and SSD was due to low sample size.

Contemporary Gene Flow

The optimal solution in BAPS was K = 3 for our badger dataset corresponding to three groups: i) Lower Peninsula of Michigan (LP), ii.) Wisconsin, Upper Peninsula of Michigan, and parts of Iowa (WI), and iii.) west of the Mississippi River (West; Figure 3.4). STRUCTURE initially detected two clusters that corresponded to the eastern United States and Canada and all other individuals (Figure S3.2). Iterative runs of the eastern cluster detected a split between the LP and all other individuals whereas the western cluster exhibited no additional substructure (Figure S3.3). The overall pattern of the three clusters with admixture within the upper Midwestern USA and Canada were similar between STRUCTURE and BAPS; disagreements on assignments largely occurred within contact zones between clusters (e.g., within Wisconsin, Minnesota, and Iowa). Therefore, the disagreement in individual assignments likely reflects either a weak barrier (Latch *et al*. 2006) or IBD because both programs can yield discrete clusters when only IBD is present (Frantz *et al*. 2009). Assignments in LP, however, were largely consistent between programs where all but three (BAPS) to five (STRUCTURE) individuals were assigned to the LP cluster.

Microsatellite diversity averaged 13.83 alleles/ locus (range: 9-20 alleles). A heterozygote deficiency was recorded in the total dataset ($F_{IS} = 0.090$, $p < 0.001$),

indicating the presence of genetic structure within our dataset. One source of structure that was evident was IBD across North America (simple Mantel test $r = 0.137$, $p < 0.001$). Furthermore, all three clusters detected in Bayesian programs (BAPS or STRUCTURE) exhibited low but significant levels of differentiation ($F_{ST} = 0.023$-0.072; all $p < 0.001$). LP was more strongly differentiated from West ($F_{ST} = 0.072$, SE = 0.021) and WI (0.052, SE = 0.012) than West versus WI (0.023, SE = 0.011). In addition to the stronger differentiation of LP from West and WI, LP also had significantly lower genetic diversity than the other two clusters (paired t-tests: all $t > 5.17$, $p < 0.001$).

LP, West, and WI clusters each exhibited heterozygote deficiencies similar to the full dataset ($F_{IS} = 0.089$-0.092, $p < 0.001$). Like the full dataset, all clusters had significant IBD according to simple Mantel tests ($r = 0.054$-0.131, all $p < 0.001$). The internal IBD within clusters combined with observed admixture between clusters, particularly between West and WI, are likely causes for the observed heterozygote deficiencies.

We recorded evidence for barrier effects in partial Mantel tests, PCA, and sPCA. Partial Mantel tests detected both Lake Michigan ($r = 0.152$, $p = 0.001$) and Mississippi River ($r = 0.065$, $p = 0.001$) as barriers to dispersal. In contrast, the Rocky Mountains and subspecific pairings were not significant (all $r < 0.023$, $p > 0.067$).

PCA also supported that Lake Michigan and Mississippi River were barriers to dispersal because groupings based on large aquatic barriers (i.e., Lake Michigan and Mississippi River) fit the data better than groupings based on subspecific definitions (Figure 3.5). Badgers east of Lake Michigan were the most genetically distinct according to the scatterplot for the first two PCA axes (Figure 3.5a). Considerable overlap in 95%

inertia ellipses were observed between those east and west of the Mississippi River, but the two groups exhibited some differentiation as compared to subspecies. Little resolution was observed among the four subspecies except for *T. t. jacksoni*, which was largely due to the presence of those east of Lake Michigan (Figure 3.5b).

Results from the sPCA were largely concordant with previous tests that showed genetic variation is dictated by both IBD and topographic barriers. The first two sPCA axes explained the most variance and had the highest eigenvalues (eigenvalues = 0.17 and 0.049; all others < 0.037), so they were considered most important in explaining genetic variation (Figure 3.6a). Both main sPCA axes exhibited high levels of spatial autocorrelation (Moran's I = 0.75 and 0.32 for axes 1 and 2 respectively; Figure 3.6b), indicating IBD. The first axis explained the most variance in the dataset (0.251) and exhibited an east-west cline with the LP having the most extreme values (Figure 3.6c). Axis 2 explained less variance than Axis 1 (0.150), but in general, individuals in the Upper Midwest were differentiated from the rest of the dataset (Figure 3.6d). Monte-Carlo tests indicated that at least one global structure (i.e., those with positive eigenvalues) observed in the sPCA was significant ($p = 0.0001$).

After forward selection, the dbRDA detected four factors that influence gene flow. The resultant model explained 10.11% of the total variation in genetic distances across North America. In the partial dbRDAs (i.e., each variable was isolated by partialling out all others), Lake Michigan explained the most variation (1.86%) in the data with Latitude as the second most important factor (1.61%). Longitude (0.91%) and Mississippi River (0.61%) were the final two significant factors (Table 3.5). None of the

landscape variables we tested (soil texture, soil order, land cover, precipitation) were significant in the partial tests.

Discussion

Badgers have strict habitat preferences that are predicted to limit gene flow outside of grassland habitats with soils suitable for digging, and to create genetic structure that is tightly associated with preferred habitats. However, badgers also exhibit high mobility and have a vast distribution that are expected to facilitate gene flow regardless of habitat and resist accumulation of spatial genetic structure. Therefore, it was unclear how grassland specialization and high mobility would interact to shape patterns of genetic variation across North America. Our data revealed broad-scale spatial genetic structure within both mitochondrial and microsatellite datasets. Differentiation was weak within mtDNA, but we recorded geographic structuring among haplotypes and a split between the Lower Peninsula, Ohio, and Indiana and the rest of the dataset. Microsatellites had more resolution than mtDNA where both individual-based and population-based approaches detected IBD and additional substructure within the Midwestern United States. Based on the geographic structuring in mtDNA and strong signatures of IBD in all microsatellite analyses, the primary influence on gene flow was geographic distance. Prominent water barriers, especially Lake Michigan, were also important in explaining patterns of genetic variation. The strongest genetic differentiation in both markers occurred between areas east and west of Lake Michigan whereas only microsatellites detected the Mississippi River as barrier to gene flow. With

IBD and water barriers being the prominent influences on gene flow, high mobility appears to be a more powerful influence than ecological specialization on broad-scale genetic patterns in badgers.

Historical Gene Flow

Mitochondrial DNA revealed high diversity, little differentiation among haplotypes, and weak geographic structuring throughout North America in badgers. Little differentiation among haplotypes and the presence of a few common haplotypes surrounded by rare haplotypes (i.e., star-like topology; Slatkin & Hudson 1991) within the haplotype network suggest that badgers did not occupy multiple, isolated glacial refugia during the LGM. Persistence in multiple glacial refugia is fairly typical for high dispersal species (e.g., Shafer *et al*. 2010; Stewart *et al*. 2010), but isolation within separate refugia typically yields well-defined groups of haplotypes that are separated by multiple mutational steps (e.g., Aubry *et al*. 2009; Latch *et al*. 2009; Lait & Burg 2013; van Els *et al*. 2014). The pattern of mitochondrial variation we observed more closely resembles patterns described for grassland species that were restricted to the Great Plains (Johnson *et al*. 2003; Oyler-McCance *et al*. 2005; Wisely *et al*. 2008; Koblmüller *et al*. 2012) or other highly mobile species that were not fragmented during the LGM (Carmichael *et al*. 2007; Teacher *et al*. 2011; Pulgarín-Restrepo & Burg 2012).

Dynamics following glacial recession in badgers appear to be more complex than other grassland specialists because the SAMOVA suggested geographic structuring in haplotypes and Quaternary fossils of badgers have been found in numerous locations outside the Great Plains (Long 1972). Our mitochondrial dataset was consistent with

badgers being somewhat widespread below the ice sheets and then multiple invasions into glaciated areas from different locations or time periods. Support for the widespread distribution of badgers during the LGM was also seen in the limited genetic differentiation, high genetic diversity and star-shaped haplotype networks because these characteristics are consistent with high effective population size (Crandall & Templeton 1993) without barriers to gene flow. The limited phylogeographic structure found in badgers greatly resembles European populations of red fox (*Vulpes vulpes*), a highly mobile, generalist carnivore thought to remained continuous during the LGM (Teacher *et al*. 2011). High gene flow in badgers, in contrast, was surprising given that grasslands underwent substantial changes during and following the LGM (Axelrod 1985; Whitlock 2000). Grasslands contracted during the late Pleistocene due to climate cooling (Axelrod 1985) and some areas (e.g., Columbia Basin and eastern Beringia; Whitlock 2000) completely shifted to forested habitats. Invasion of forests into grasslands did not necessarily mean range contraction for badgers, however, because many areas maintained mixes of forest and grasslands that are still suitable habitats for badgers (e.g., Duquette & Gehrt 2014). Thus, the variable and heterogeneous habitats below the ice sheets may have remained suitable for badgers, allowing for broad-scale movement and retention of genetic diversity.

Despite the dynamic history of grasslands and numerous topographic barriers in North America, the SAMOVA only split the dataset into samples east and west of the Lake Michigan. Western populations appear to have expanded north from southern areas such as Utah and the Great Plains due to strong signatures of population expansion in demographic tests and high diversity below ice sheets. Expanded populations, in

contrast, typically exhibit decreased genetic diversities and persistence of common haplotypes at high frequencies due to repeated founder effects (e.g., Hewitt 2000). Ethier *et al*. (2012) recorded much more substructure in northern latitudes than this study, which would occur if northern populations experienced bottlenecks associated with colonization from southern populations. The decreased genetic diversity and high rates of drift would create higher genetic differentiation even without physical barriers to gene flow. When we included southern localities that had substantially higher diversity, we did not find any discrete groups despite similar geographic structuring of Haplotypes 2 and 4. Therefore, our data indicates that most of western North America maintained high effective population sizes and gene flow that prevented strong differentiation in mtDNA, and founder effects occurred in northern latitudes resulting in high frequencies of Haplotype 2.

Mechanisms shaping geographic structuring of haplotypes within the Midwest were apparent, with populations exhibiting clear signatures of one or more population expansions. Previously glaciated regions west of Lake Michigan (e.g., Wisconsin and Minnesota) exhibited high frequencies of Haplotype 1, consistent with expansion from the Great Plains. The presence of Haplotype 3 in Ohio and the LP, in contrast, suggests two potential colonization scenarios for the LP. First, badgers could have colonized eastern North America once and Lake Michigan could have isolated badgers in the LP, resulting in the split between eastern (i.e., LP, Ohio, and Indiana) and western badgers observed in the SAMOVA. Haplotype 3 is one transition away from Haplotype 1, the haplotype most common in surrounding states (i.e., Minnesota, Wisconsin, and Iowa), so Haplotype 3 could have arisen post colonization and then Lake Michigan prevented the

spread of Haplotype 3 back into western North America. Alternatively, a separate colonization in the LP caused the high frequency of Haplotype 3 in the LP. Guilday (1968) suggested that eastern North America was invaded twice by prairie-associated species based on fossils in both pre-and post-Wisconsin deposits in Kentucky, Maryland, and Pennsylvania. Multiple colonization routes into the Great Lakes region has been recorded in herpetofauna (Austin *et al*. 2002; Zamudio & Savage 2003; Placykr Jr *et al*. 2007), resulting in disparate frequencies of haplotypes between the LP and surrounding states much like Haplotype 3 in this study. The few mammal studies in the Great Lakes region are largely limited to forest-associated species (Rowe *et al*. 2004; Taylor & Hoffman 2010; Reding *et al*. 2012), recent invaders (Koblmüller *et al*. 2012), or those not found in the Lower Peninsula (Koblmüller *et al*. 2009; vonHoldt *et al*. 2011), so it is difficult to determine if multiple colonization routes also occurred in mammals including badgers. Both mechanisms, a founder effect followed by isolation or two colonization routes, would result in the high frequency of Haplotype 3 and the corresponding split between eastern (i.e., LP, Ohio, and Indiana) and western North America. Therefore, further studies could help differentiate between these two mechanisms via hypothesis testing of separate recolonization scenarios (e.g., Frantz *et al*. 2014) with alternative markers that allow divergence time estimation.

Contemporary Gene Flow

Grasslands have undergone massive changes following the LGM including widespread conversion to agriculture. Conversion to agriculture has reduced the amount of grasslands and increased fragmentation among grassland patches, resulting in

demographic declines and reduced movement between remaining patches in many grassland-associated taxa (e.g., Coppedge *et al*. 2001; Steffan-Dewenter & Tscharntke 2002; Wisely *et al*. 2008). Genetic differentiation can occur if individuals cannot reach distant grassland patches, and resultant genetic variation would be correlated with the presence of grassland habitat. Our microsatellite dataset did not follow this expectation because we detected both IBD and topographic barriers only in all analyses. Instead, patterns of genetic structure within our microsatellite markers were remarkably similar to our mitochondrial dataset (i.e., geographic structuring and strong divergence within the Great Lakes region). Therefore, high dispersal appears to be the primary force in driving patterns of genetic variation across North America.

The overall lack of structuring according to ecological factors in badgers suggests that their high mobility counteracts any isolating impacts of habitat heterogeneity, particularly in western North America. Much of our dataset was contained within the West cluster, a single genetic population characterized by IBD. Mobile specialists can experience extensive connectivity in suitable habitat, particularly in the absence of topographic barriers (e.g., Carmichael *et al*. 2007; Marthinsen *et al*. 2009; Row *et al*. 2012). However, the large West cluster in badgers was surprising because it encompassed many potentially isolating barriers, including those that defined population boundaries in previous studies (e.g., mountain ranges; Kyle *et al*. 2004; Ethier *et al*. 2012). We likely found limited genetic subdivision in western North America because detection of genetic differentiation heavily depends on the balance between genetic drift and gene flow, two evolutionary processes that promote and resist genetic differentiation respectively. The rate of genetic drift in a population is directly correlated with effective

population size (Wright 1931), and previous studies included endangered Canadian populations whose total census sizes are 200 or less (COSEWIC 2012). Even in these small populations, mountain ranges were not complete barriers to gene flow (Kyle *et al*. 2004; Ethier *et al*. 2012), so the extensive gene flow within western North America appears to overcome any isolating effects of the Rocky Mountains or ecological variables. Therefore, badgers appear to maintain large population sizes outside isolated, peripheral areas, which combined with their high mobility resists genetic differentiation.

Badgers in eastern North America formed two genetic clusters and had lower genetic diversity than western areas, which at least in part result from large water barriers (i.e., Mississippi River and Lake Michigan) preventing gene flow. Both the Mississippi River and Lake Michigan appear to disrupt enough gene flow to allow genetic drift to cause detectable genetic differentiation. Considerable admixture and disagreement between Bayesian clustering programs was observed between the West and Wisconsin clusters, so greater gene flow likely occurs across the Mississippi River than Lake Michigan. Another factor that may enhance genetic differentiation observed in eastern North America is that eastern badgers occur at a range periphery. Genetic drift is stronger in populations with small effective populations sizes, and effective population sizes in peripheral populations can be much lower than core areas (Vucetich & Waite 2003). Eastern North America certainly contains less grassland habitat than western North America, which may cause lower population sizes and high fluctuations typical of peripheral populations (Hengeveld & Haeck 1982). Based on the high habitat heterogeneity at the eastern periphery, genetic differentiation due to ecological variables should have been detectable if gene flow was restricted as observed in other highly

mobile species (e.g., Schwartz *et al*. 2009; Garroway *et al*. 2011; Weckworth *et al*. 2013; Balkenhol *et al*. 2014). However, we only detected genetic differentiation according to topographic barriers, so badgers appear able to maintain high gene flow except when faced with large geographic barriers.

Lake Michigan prevents gene flow from western populations (i.e., Upper Peninsula and Wisconsin) as evidenced by the strong assignments of most LP individuals to the LP cluster. The LP, therefore, mainly receives migrants from southern areas that have low density badger populations (e.g., Warner & Ver Steeg 1995; Duquette & Gehrt 2014). A lack of connectivity with the genetically diverse populations to the west facilitated the loss of genetic diversity in LP badgers following postglacial recolonization. Other peninsular populations with low genetic diversity have been recorded in highly mobile species, and most have found genetic variation patterns consistent with a combination of historic founder events and subsequent isolation within a peninsula (Tammeleht *et al*. 2010; Reding *et al*. 2012; Frantz *et al*. 2014). Our data supports a historic founder event and subsequent isolation because no demographic declines have been recorded in the LP and both genetic datasets exhibited low diversity. Very little life history information is available for badgers in the LP, but previous genetic work in the LP suggested three potential mechanisms for heterozygote deficiencies: Wahlund Effect, mixing of populations, and inbreeding (Kierepka *et al*. 2012). We also recorded a heterozygote deficiency within the LP, but with the strong assignments to a single cluster in both Bayesian programs, badgers in the LP appear to experience some degree of inbreeding. Further assessment of the LP should focus on determining if badgers in the LP warrant conservation actions like other peripheral populations in British Columbia

and Ontario. The potential for inbreeding as a mechanism reducing genetic variation in LP badgers, and other peripheral populations, should be further explored by pairing it with demographic and fitness related data to assess any future conservation needs.

Conclusions

Our broad-scale genetic investigation of gene flow in badgers revealed limited genetic structure across much of North America. IBD largely defined genetic structure within microsatellites, but we still detected signatures of historical biogeographic events. Badgers in western North America appear to maintain a robust population size based on the high genetic diversity and limited differentiation west of the Mississippi River. In contrast, eastern North America had genetic differentiation due to the Mississippi River and Lake Michigan as well as lower genetic diversity, so we believe eastern badgers have lower or more variable population sizes than western grassland-dominated areas. In particular, the LP was genetically distinct in all analyses due to both low genetic diversity and barrier effects of the Lake Michigan, which suggests a combination of recolonization following glacial retreat and peninsular geography driving genetic variation. Taken together, high mobility appears to prevent genetic differentiation according ecological factors in badgers, but differences in genetic diversity and differentiation between western and eastern North America may reflect variation in habitat quality impacting the strength of genetic drift. Therefore, even if high mobility in badgers is the dominant driver of gene flow across North America, their specialist traits may influence genetic variation through differences in effective population sizes according to habitat quality.

With limited genetic structure in badgers observed throughout much of North America, patterns of mitochondrial and microsatellite variation conflict with all current morphological subspecific designations. Previous studies have found evidence for a split between *T. t. taxus* and *jeffersonii* in peripheral, endangered populations (Kyle *et al*. 2004; Ethier *et al*. 2012), but with the inclusion of more southern locations in this study, no such divergence was found. Discordance between morphological subspecies and genetic variation in North America is relatively common (e.g., Zink 2004; Cullingham *et al*. 2008; Godbout *et al*. 2008; Hull *et al*. 2008; Sabatino & Routman 2009; Kodandaramaiah *et al*. 2012; Latch *et al*. 2014), which may argue for taxonomic revision of badgers. However, proper delineation of subspecies should consider multiple datasets including genetics, morphology, and behavior (Haig *et al*. 2006), and our dataset based on neutral genetic markers does not provide a complete picture of how subspecific boundaries may reflect evolutionary divergence. In particular, neutral datasets may not reflect variation at adaptive loci, and morphological differences could reflect selection pressures that coincide with subspecific boundaries. Therefore, any decision to re-evaluate taxonomic classification should incorporate potential adaptive drivers of morphological differentiation in badgers, especially since no common thread appears to exist in how carnivores react to landscape heterogeneity.

Badgers join the growing number of highly mobile species that demonstrate the difficulty in predicting how dispersal capability and habitat preferences impact gene flow in widespread taxa. Badgers represent somewhat of a unique case among specialists because we observed limited genetic differentiation across multiple habitat types, a finding that usually occurs only within suitable habitat (Carmichael *et al*. 2007; McRae &

Beier 2007; Koen *et al*. 2012; Row *et al*. 2012). Therefore, reliance on grasslands and fossorial rodents does not preclude badgers from modified or sub-optimal habitats, but maintaining corridors (e.g., remnant habitats in high agricultural landscapes; Duquette & Gehrt 2014) may be critical for continued high gene flow within peripheral populations. In conclusion, this broad-scale genetic survey of badgers revealed that dispersal capabilities and specialization influence both gene flow and effective population size, and thus, can produce disparate genetic patterns across a large geographic range.

Figure 3.1. Sampling locations and subspecific designations of $n = 917$ badgers. Designations are morphological subspecies as defined in Long (1972).

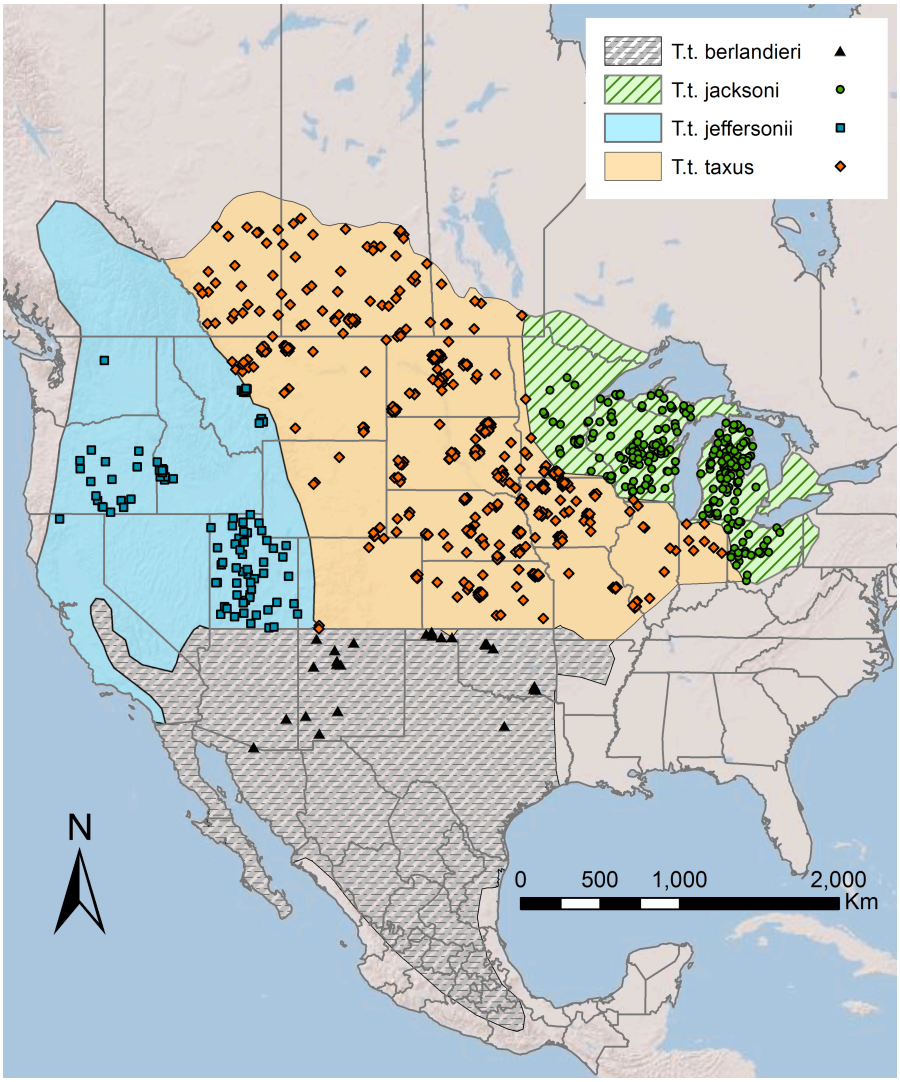
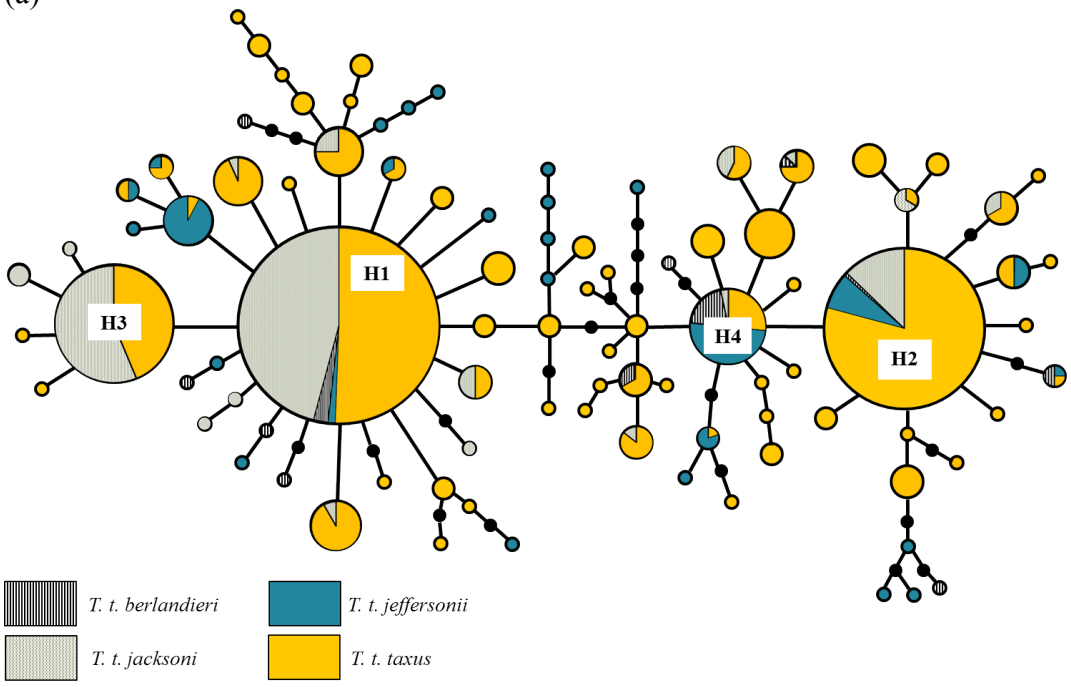


Figure 3.2. Haplotype network for 95 d-loop haplotypes detected in *n* = 785 badgers. Each circle represents a single haplotype and the size of the circle corresponds to its frequency. Each line represents a single mutational step, and black circles indicate missing genotypes. The network was color-coded by the morphological subspecies (a) or inferred subspecies (b). The most common haplotypes (H1, H2, H3, and H4) are labeled.

(a)



(b.)

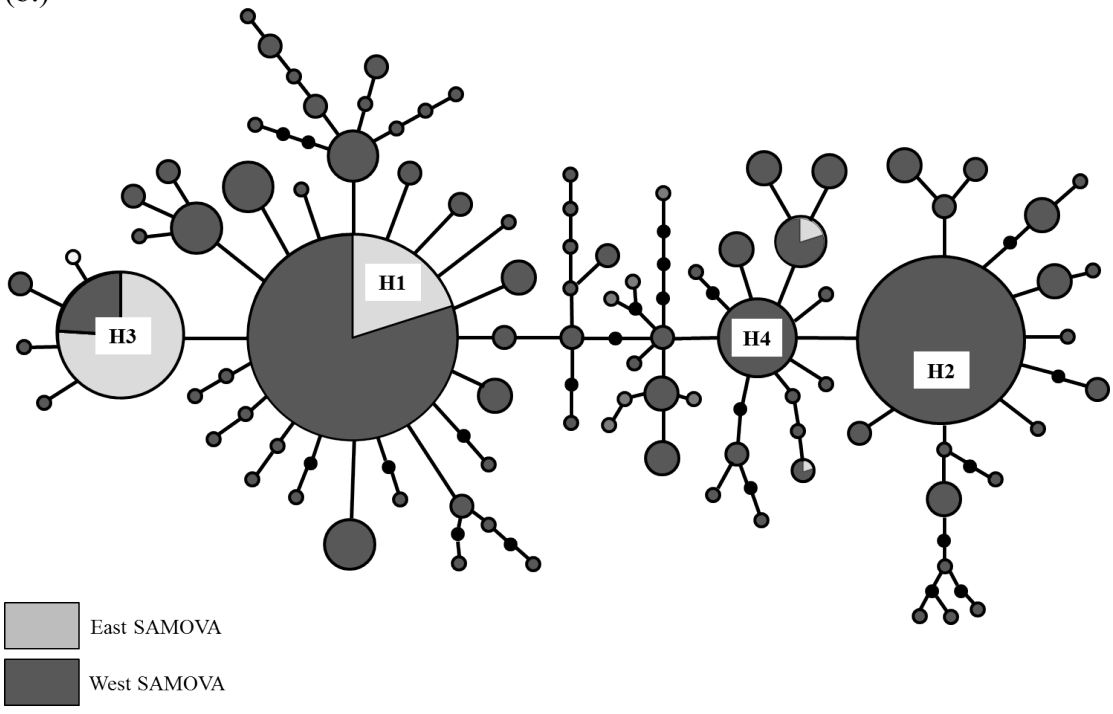


Figure 3.3. Mismatch distributions for the total dataset (a), subspecies (b-e), and SAMOVA groups (f-g). Observed distributions are represented as gray bars and expected distributions under rapid demographic expansion are black solid lines. 95% confidence intervals for the expected distribution are given as dotted black lines. All groups followed expected patterns under rapid demographic expansion (all $p < 0.008$) except for the eastern SAMOVA group ($p < 0.002$).

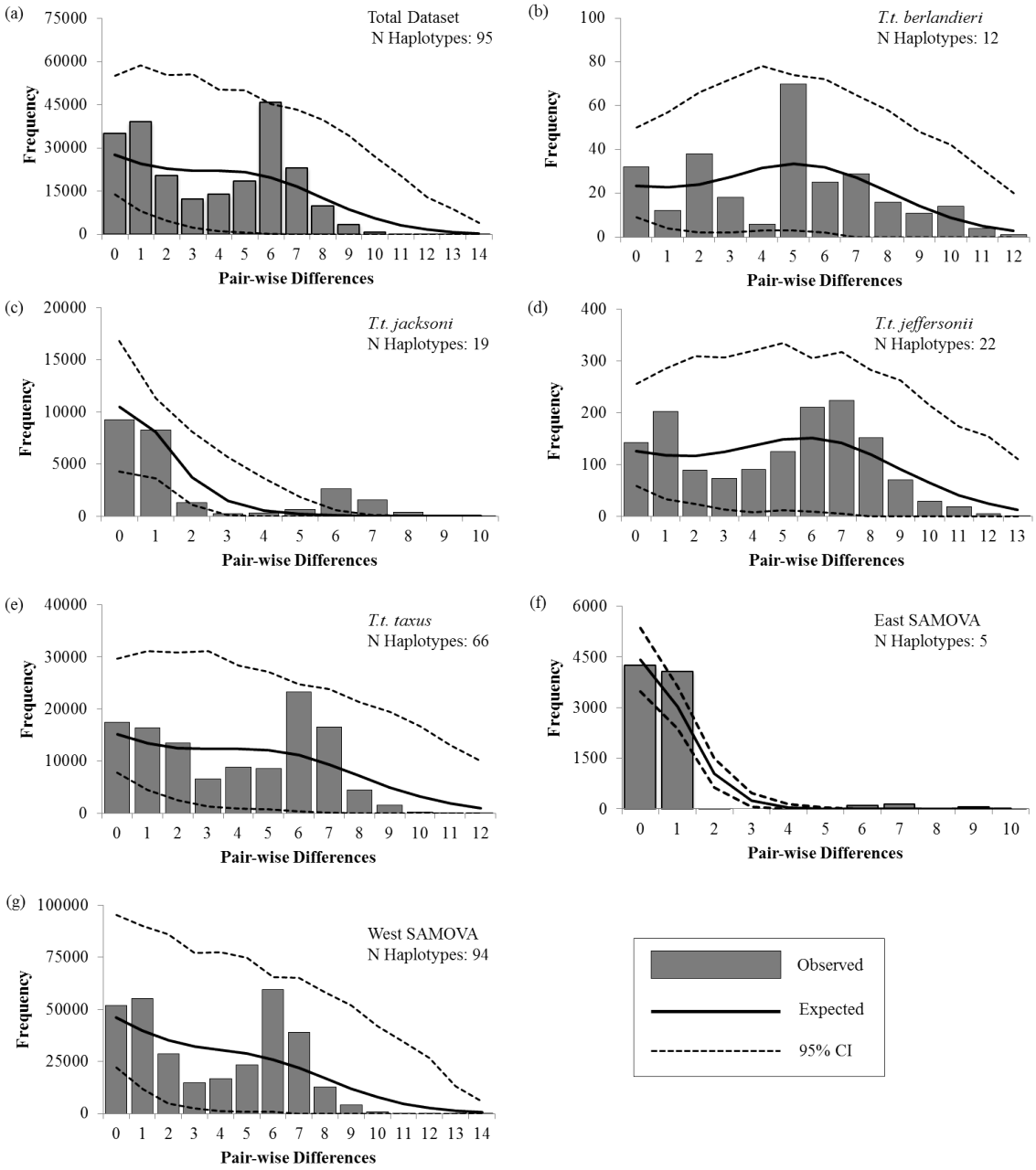


Figure 3.4. Distribution of three genetic clusters inferred in BAPS 5. The three clusters (West: black; Wisconsin: grey, and LP: white) are largely separated by two major water barriers: the Lake Michigan and the Mississippi River. The Mississippi River is represented by the bold line.

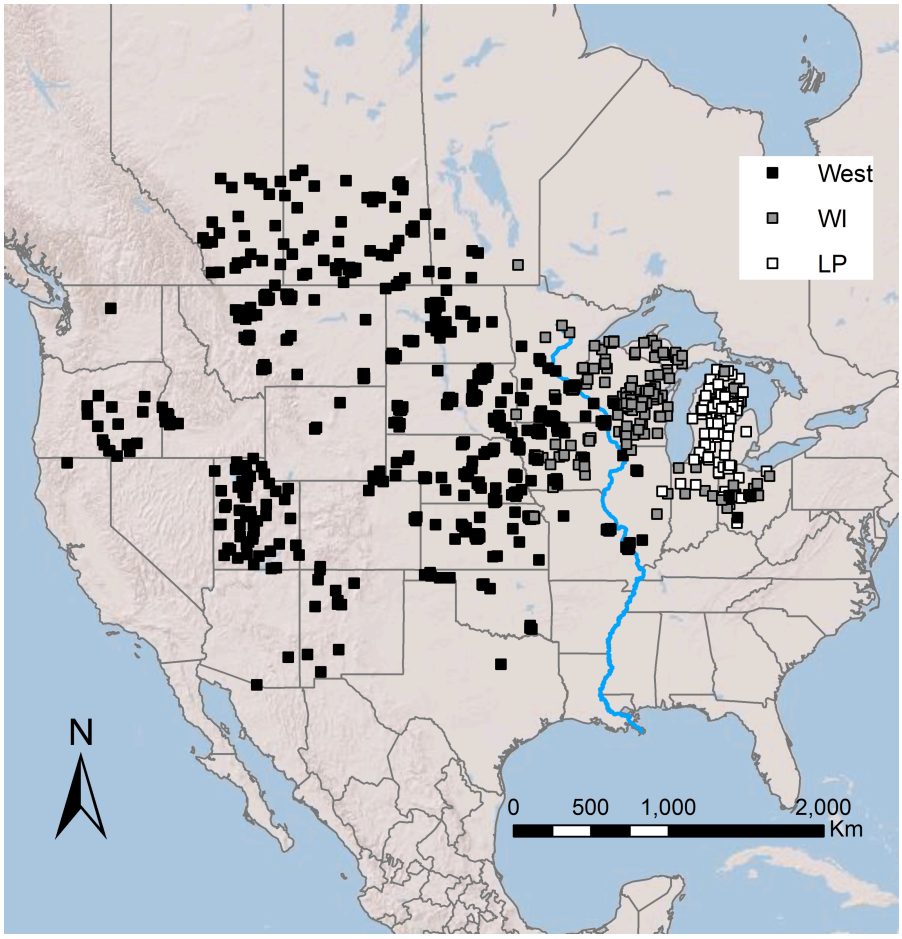
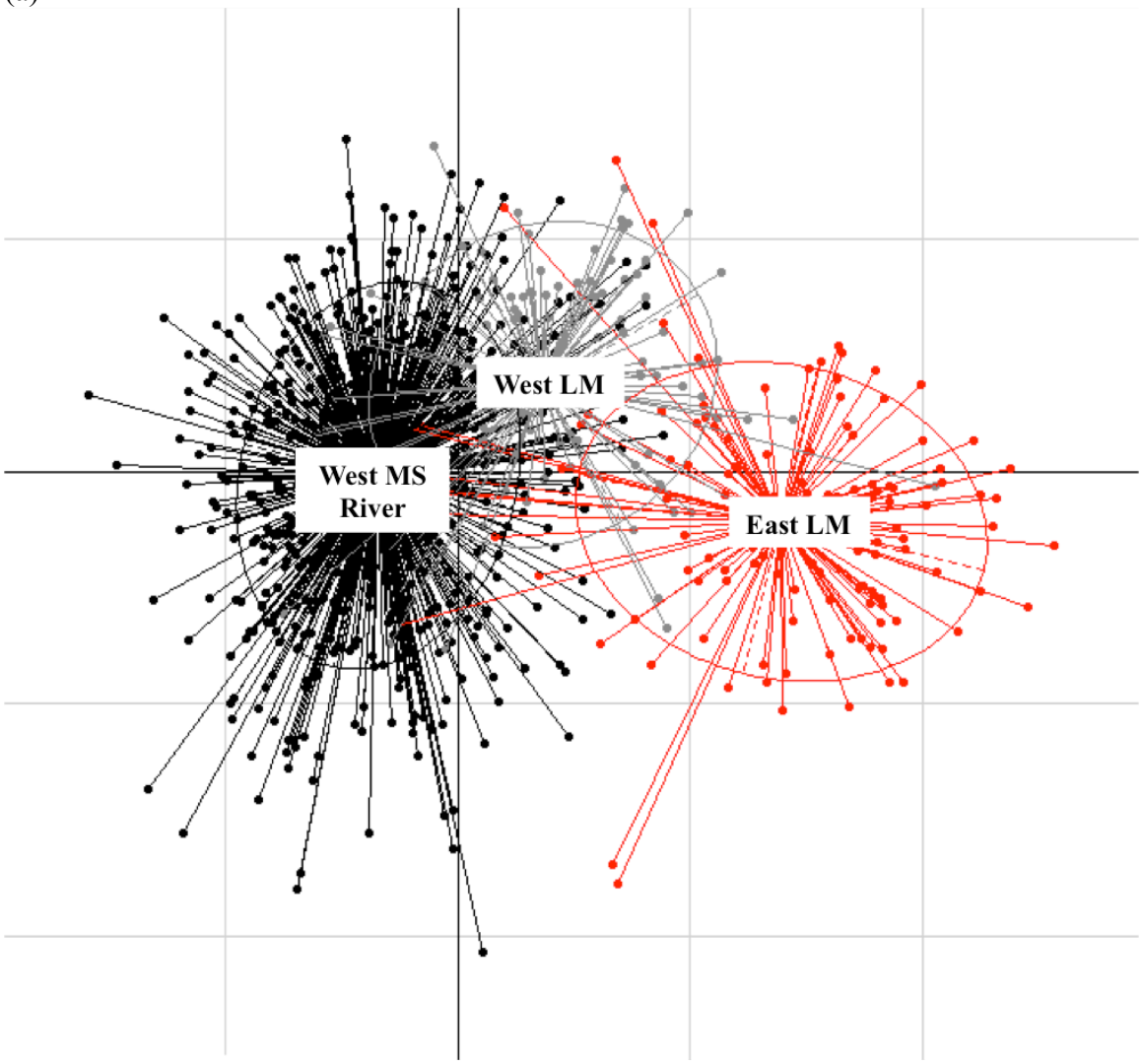


Figure 3.5. Plot of first two PCA axes when grouped by large aquatic barriers (a) or subspecies (b). Each point represents a single badger's genotype that is color-coded according to either the aquatic barriers (east of Lake Michigan, west of Lake Michigan/East of Mississippi River, and west of the Mississippi River) or subspecies (*T. t. berlandieri*, *jacksoni*, *jeffersonii*, or *taxus*). Ovals around points correspond to 95% confidence ellipses that denote where 95% of individuals within each group occur. Each square on the background grid corresponds to 0.5.

(a)



(b)

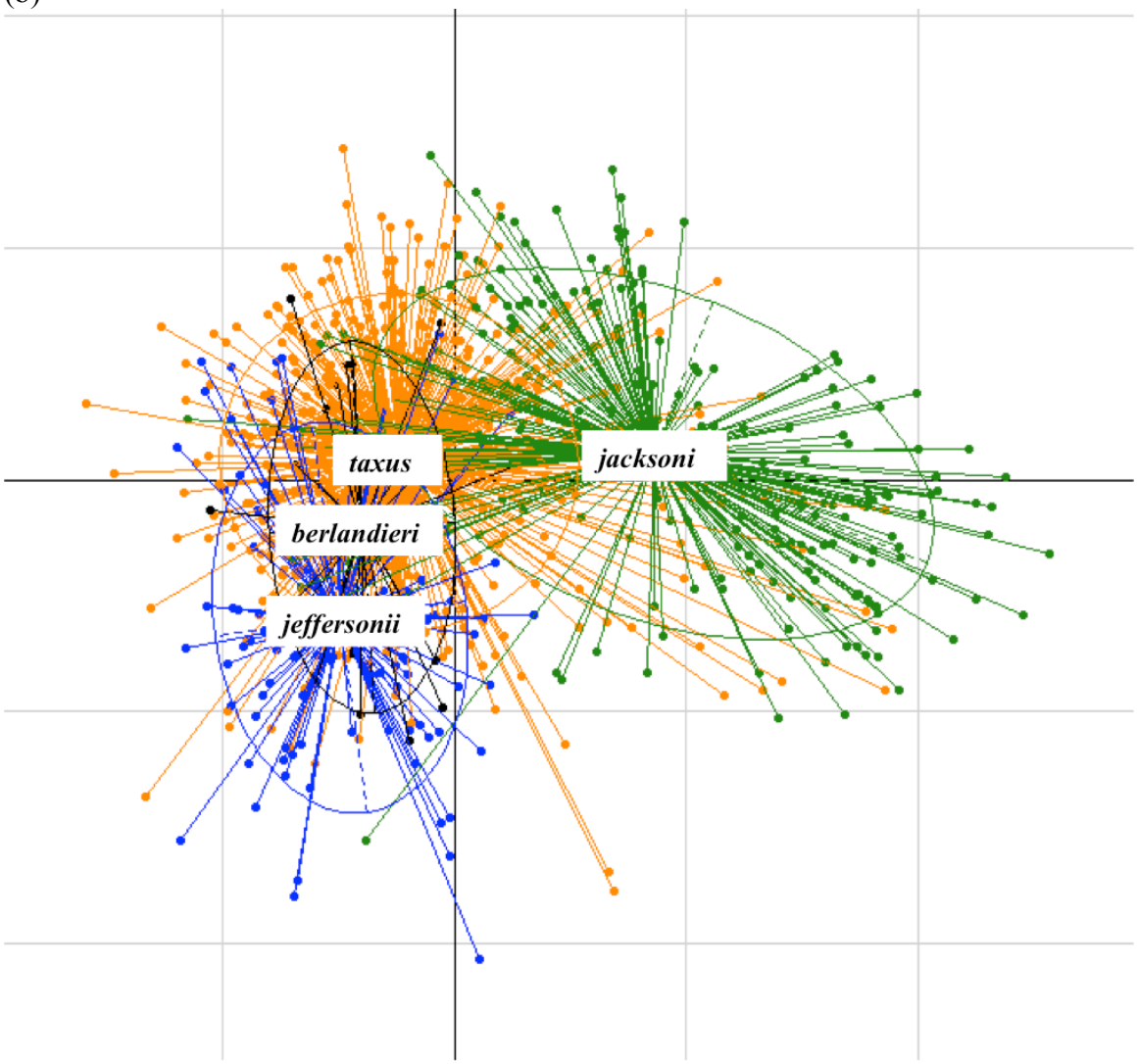


Figure 3.6. Results of the sPCA analysis. Eigenvalue plots (a) and the screeplot (b) indicate that the first two axes are most important in explaining genetic variation. The screeplot plots the relationship between variance explained by each axis and spatial autocorrelation (Moran's I) within each axis. Spatially lagged scores from Axes 1 (c) and 2 (d) depict two distinct patterns in genetic variation. Squares represent each individual where larger squares have stronger positive (white) or negative (black) values whereas smaller, grey (light grey = positive, dark grey = negative) squares are less strongly differentiated.

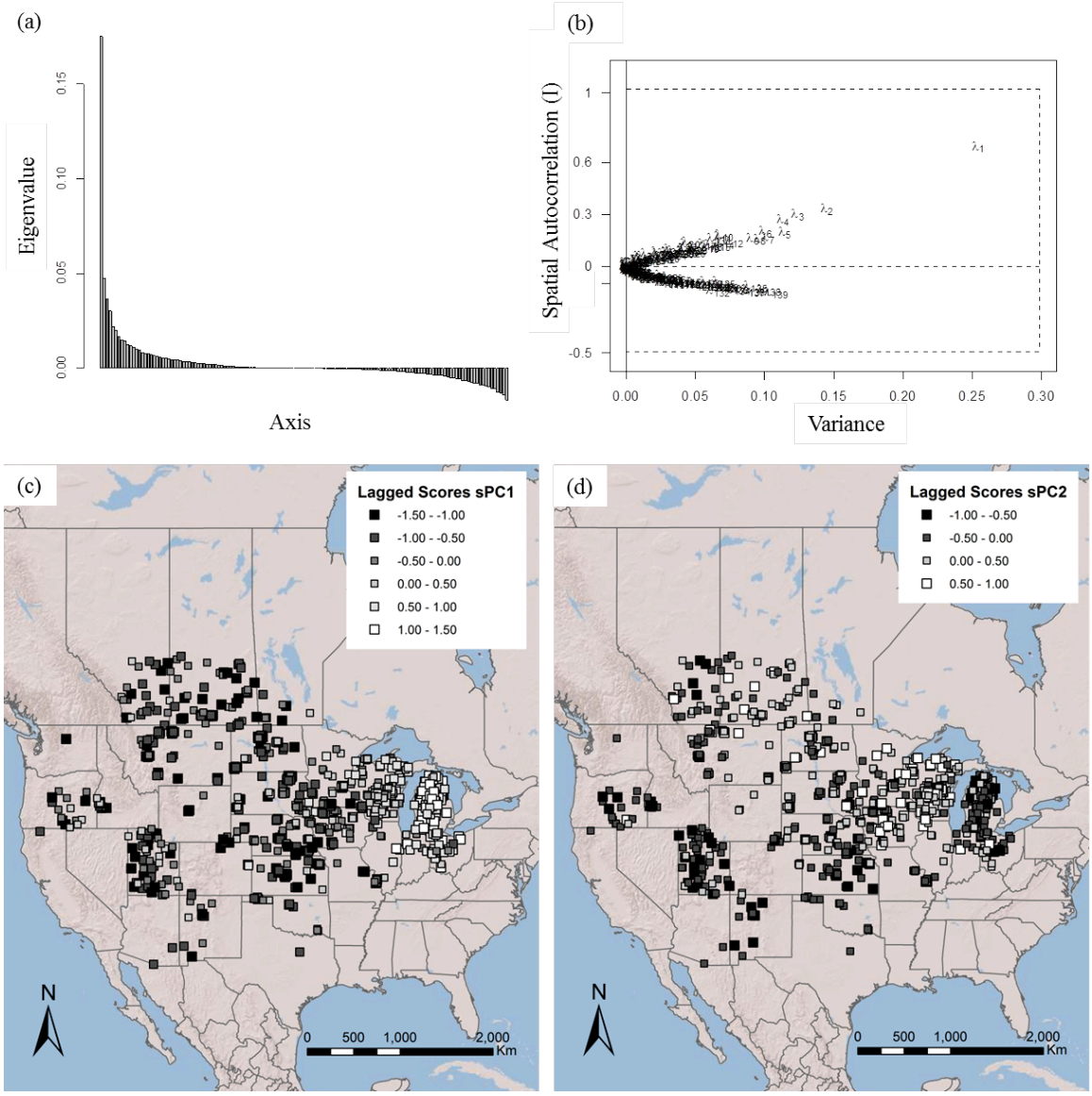


Figure S3.1. Geographic distribution of the four main haplotypes and those with the 26 bp deletion region that were not included within the mtDNA analyses. Rare haplotypes had a frequency of 10 or less individuals (dark grey circles). Haplotype 1 (red squres) is found throughout North America, but its frequency is the highest in the Upper Midwest. Haplotype 2 (blue circles) primarily occurs within northern latitudes, particularly in Montana, Canada, and North Dakota. Haplotype 3 (yellow diamonds) is one base pair away from Haplotype 1, and is chiefly located in the Lower Peninsula of Michigan and Ohio. Haplotype 4 (light blue triangles) is one mutation away from Haplotype 2, and almost always occurred west of the Rocky Mountains. Like Haplotype 4, individuals with the 26 bp deletion region (green squares) are primarily found west of the Rocky Mountains. The haplotype network of the 95 haplotypes that did not include the deletion region is provided where the four main haplotypes are colored.

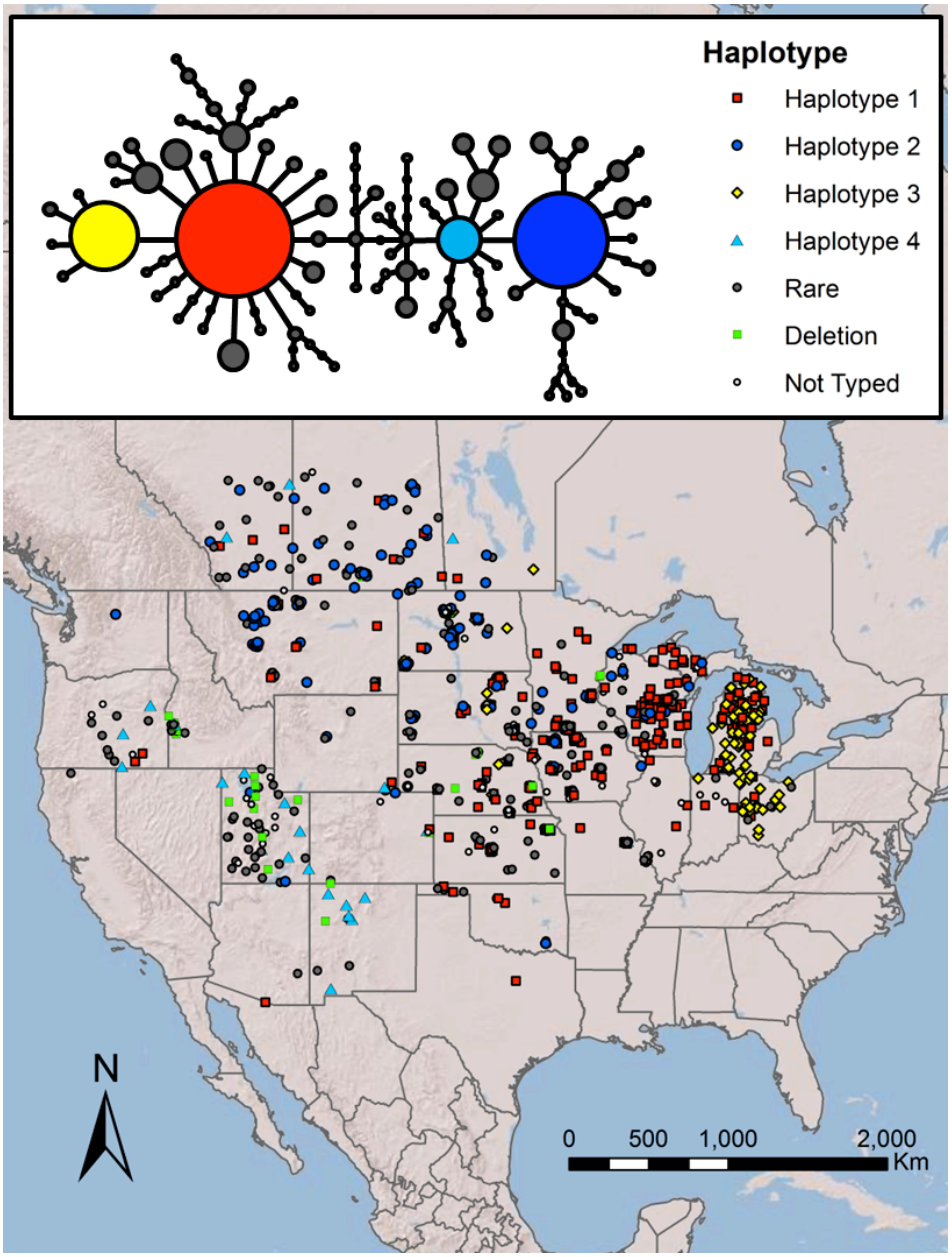
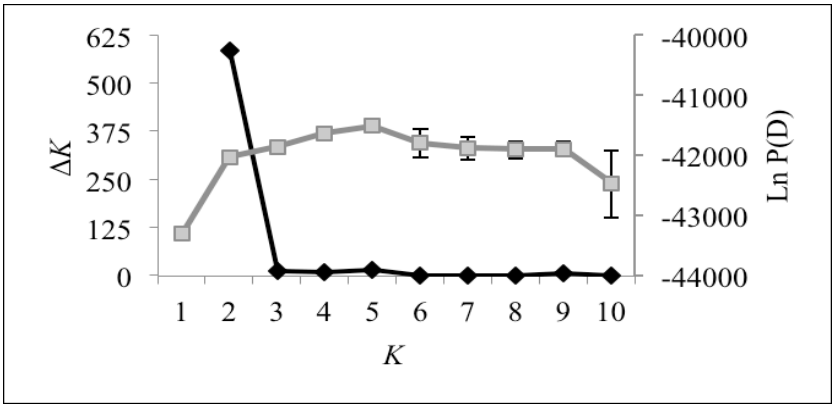


Figure S3.2. Results of the overall clustering analysis within STRUCTURE. Δ*K* indicated the most likely number of clusters was 2 as shown by the large peak at *K* = 2 within the plot (a). The spatial arrangement of q-values (b) indicated a general cline from east to west. We interpolated *q*-values using inverse-weight distance to create a genetic surface across our sampling area and individuals are colored according to their q-values. Those with higher *q*-values (large blue circles) were more strongly assigned to the eastern cluster.

(a)



(b)

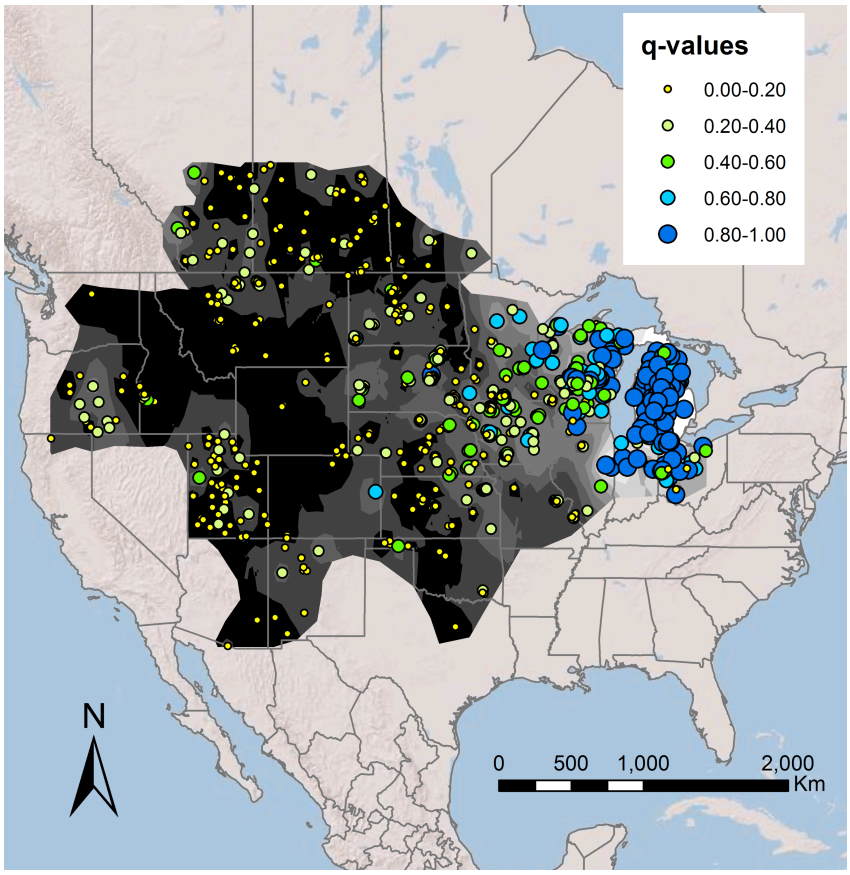
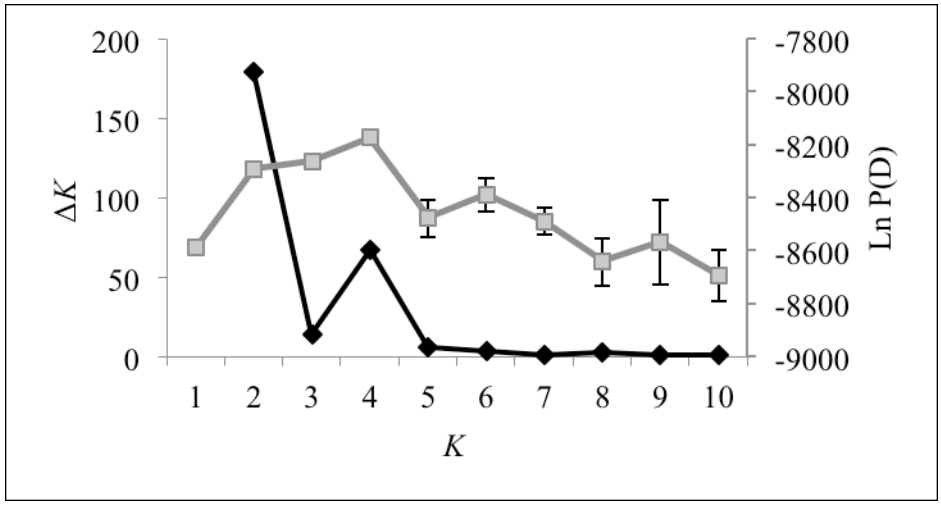
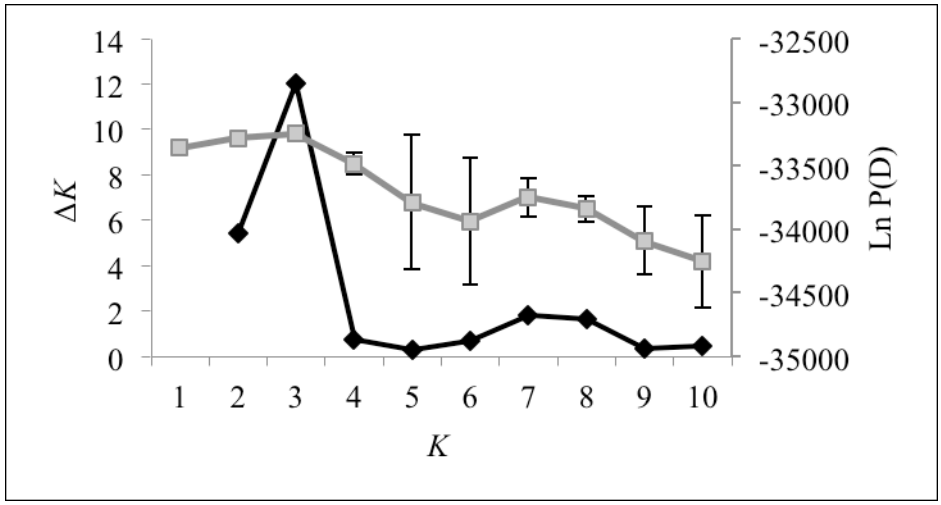


Figure S3.3. Results of the iterative runs of the two main clusters detected in the overall clustering analysis within STRUCTURE. The Δ*K* plot for the eastern cluster (a) indicated further substructure whereas the plot for the western cluster (b) did not strongly support any additional genetic substructure. When mapped (c), the iterative run in the eastern cluster (n = 210 individuals) suggested the presence of two additional clusters that separated the Lower Peninsula of Michigan from the rest of the individuals. Interpolated *q*-values, created via inverse-weight distance interpolation, revealed a strong divergence between the Lower Peninsula and Wisconsin (high assignment to the Lower Peninsula is denoted with large blue circles).

(a)



(b)



(c)

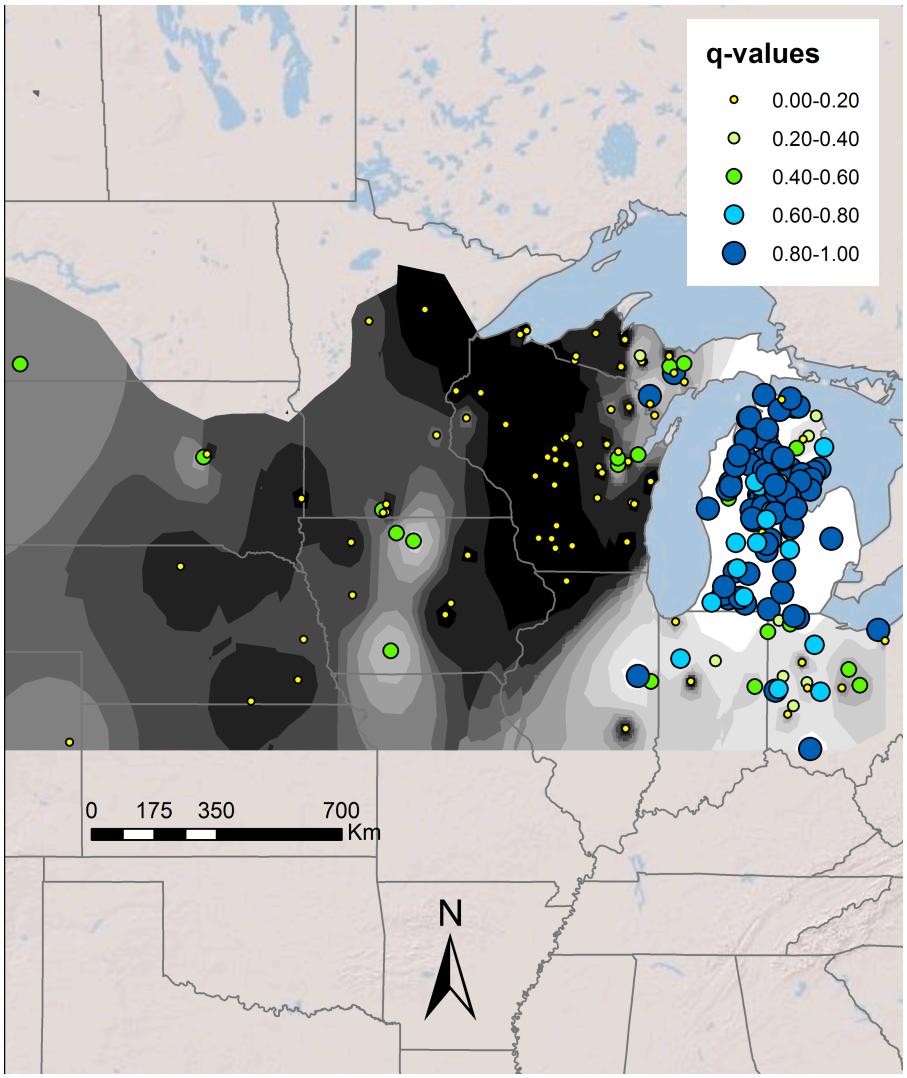

q-values
0.00-0.20
0.20-0.40
0.40-0.60
0.60-0.80
0.80-1.00
0
175
350
700
Km
N

Table 3.1. Results of SAMOVA for K = 2 where the first group contained the Lower Peninsula of Michigan, Indiana, and Ohio and the second group contained all other individuals. All variance components were significant after 1000 permutations (all $p < 0.001$).

Source	Df	SS	Components	% Var
Among Groups	1	138.955	0.564	25.08
Among Pops Within Groups	17	129.667	0.161	7.18
Within Pops	765	1164.412	1.522	67.74
Total	783	1433.04	2.247	100.00

Table 3.2. Diversity metrics for n = 785 badgers, for each of the four subspecies, and for the two SAMOVA groups. All diversity measures include standard errors.

Grouping	N	*H*	π	Pairwise diff
Total	785	0.831 ± 0.099	0.006 ± 0.003	3.660 ± 1.855
Subspecies				
T. t. berlandieri	24	0.873 ± 0.047	0.008 ± 0.005	4.681 ± 2.376
T. t. jacksoni	222	0.325 ± 0.034	0.003 ± 0.002	1.856 ± 1.068
T. t. jeffersonii	54	0.926 ± 0.040	0.008 ± 0.005	4.774 ± 2.370
T. t. taxus	485	0.786 ± 0.045	0.007 ± 0.004	3.827 ± 1.928
SAMOVA grouping				
East SAMOVA	134	0.523 ± 0.023	0.001 ± 0.001	0.800 ± 0.582
West SAMOVA	651	0.813 ± 0.011	0.007 ± 0.004	3.319 ± 1.924

N: Sample sizes for each group; *h*: haplotype diversity; π: nucleotide diversity

Table 3.3. Mismatch statistics for the four subspecies and for the two SAMOVA groups. Fu's Fs and Tajima's D indicate demographic expansion when they are significantly negative whereas significant SSD and raggedness indices suggest deviations from patterns expected under demographic expansion. P-values for each test are given in parentheses; significant values ($P < 0.05$) are in bold.

Group	*N*	*Fu's Fs*	*Tajima's D*	*SSD*	*Raggedness*
Total	785	**-25.075**	**-1.352**	0.036	0.025
		(0.001)	**(0.015)**	(0.422)	(0.112)
T. t. berlandieri	24	**-25.464**	-0.516	0.106	0.033
		(0.001)	(0.328)	(0.125)	(0.423)
T. t. jacksoni	222	**-27.350**	-0.895	0.096	0.031
		(0.001)	(0.125)	(0.098)	(0.282)
T. t. jeffersonii	54	**-25.618**	-0.729	0.020	0.014
		(0.001)	(0.221)	(0.324)	(0.428)
T. t. taxus	485	**-25.220**	-1.005	0.035	0.022
		(0.001)	(0.120)	(0.183)	(0.212)
East SAMOVA	134	0.048	-1.384	**0.029**	**0.211**
		(0.563)	(0.065)	**(0.001)**	**(0.002)**
West SAMOVA	651	**-25.096**	-1.346	0.023	0.038
		(0.001)	(0.059)	(0.288)	(0.346)

Table 3.4. Genetic diversity measures for three genetic clusters detected in BAPS. All three populations had significant heterozygote deficiencies (bold), and LP had lower genetic diversity than WI and West (all $t > 2.341$, $p < 0.02$). P-values are given in parentheses for F_{IS} values.

	N	A_R	H_O	H_E	F_{IS}
LP	115	7.99	0.641	0.705	**0.089 (0.001)**
WI	155	10.22	0.731	0.804	**0.092 (0.001)**
West	649	11.16	0.747	0.816	**0.086 (0.001)**

A_R: allelic richness corrected for smallest sample size of 115; H_O: observed heterozygosity; H_E: expected heterozygosity; F_{IS}: Weir & Cockerham (1984)'s inbreeding coefficient

Table 3.5. Results from partial dbRDA analyses to evaluate the relationship between genetic distance and topographic barriers, geographic coordinates, and landscape variables. The final model after sequential selection included Latitude, Longitude, Lake Michigan, and the Mississippi River (bold), and together explained 10.11% of the total genetic variation in the dataset. Variance explained by each singular variable with all other variables partialled out are provided (% var).

Variable	F-ratio	p-value	% var
Latitude	**16.31**	**0.0001***	**1.41**
Longitude	**10.54**	**0.0001***	**0.91**
Lake Michigan	**21.45**	**0.0001***	**1.86**
Mississippi River	**7.92**	**0.0002***	**0.69**
Rocky Mountains	3.95	0.0260*	0.34
Subspecies	1.86	0.1500	0.16
Soil Order	5.09	0.0240*	0.44
Soil Texture	0.37	0.9020	0.09
Precipitation	4.12	0.0300*	0.41
Land Cover	3.25	0.0510	0.22

*corrected alpha = 0.015

Table S3.1. Information about all the badgers, location, and source. Location corresponds to the state where badgers were sampled and for those that had small sample sizes (< 10 individuals), their SAMOVA group is listed in parentheses. All other SAMOVA groups are identical to their capture state or Canada. Sample sizes (N) correspond to the number of individuals provided by each collector and organization.

Location	N	Organization
Alberta (Canada)	5	North American Fur Auction
Arizona (New Mexico)	2	Arizona Fish and Game Department
California (Oregon)	1	Oregon Department of Fish and Wildlife
Canada	46	North American Fur Auction
Colorado (Wyoming)	5	North American Fur Auction
Colorado (New Mexico)	1	Museum of Southwestern Biology
Iowa	63	North American Fur Auction
Idaho	20	Idaho Fish and Game
Idaho	1	North American Fur Auction
Illinois	13	North American Fur Auction
Illinois	1	Illinois Department of Natural Resources
Indiana (Ohio)	5	The Ohio State University
Indiana (Ohio)	2	North American Fur Auction
Kansas	1	Prairie Wildlife Research
Kansas	42	North American Fur Auction
Lower Peninsula, Michigan	95	Central Michigan University, Michigan Department of Natural Resources
Lower Peninsula, Michigan	4	North American Fur Auction
Manitoba (Canada)	4	North American Fur Auction
Minnesota	70	North American Fur Auction
Missouri (Iowa)	8	North American Fur Auction
Montana	6	Oregon Department of Fish and Wildlife
Montana	51	North American Fur Auction
Nebraska	65	North American Fur Auction
New Mexico	10	Museum of Southwestern Biology
New Mexico	1	North American Fur Auction
North Dakota	64	North American Fur Auction
Ohio	29	The Ohio State University
Ohio	1	North American Fur Auction
Oklahoma	9	Oklahoma Department of Wildlife Conservation
Oklahoma	3	North American Fur Auction
Oregon	16	Oregon Department of Fish and Wildlife
Oregon	1	North American Fur Auction
Saskatchewan (Canada)	26	North American Fur Auction
South Dakota	87	North American Fur Auction
South Dakota	1	Prairie Wildlife Research
Texas (Oklahoma)	1	Oklahoma Department of Wildlife Conservation

Upper Peninsula, Michigan	18	Central Michigan University, Michigan Department of Natural Resources
Utah	50	Utah Trappers Association
Utah	3	North American Fur Auction
Washington (Oregon)	1	Museum of Southwestern Biology
Wisconsin	1	North American Fur Auction
Wisconsin	76	University of Wisconsin-Milwaukee, Wisconsin Department of Natural Resources
Wyoming	1	Prairie Wildlife Research
Wyoming	7	North American Fur Auction

Literature Cited

Anderson MJ (2003) DISTLM Forward: A FORTRAN Computer Program to Calculate a Distance-based Multivariate Analysis for a Linear Model Using Forward Selection. Department of Statistics,University of Auckland, New Zealand.

Anderson MJ (2004) DISTLM v.5: A FORTRAN Computer Program to Calculate a Distance-based Multivariate Analysis for a Linear Model. Department of Statistics, University of Auckland, New Zealand.

Aubry KB, Statham MJ, Sacks BN, Perrine JD, Wisely SM (2009) Phylogeography of the North American red fox: vicariance in Pleistocene forest refugia. *Molecular Ecology*, **18**, 2668-2686.

Austin JD, Lougheed SC, Neidrauer L, Chek AA, Boag PT (2002) Cryptic lineages in a small frog: the post-glacial history of the spring peeper, *Pseudacris crucifer* (Anura: Hylidae). *Molecular Phylogenetics and Evolution*, **25**, 316–329.

Axelroad DI (1985) Rise of the grassland biome, central North America. *Botantical Review*, **51**, 163-201.

Azevedo FCC, Lester V, Gorsuch W, Larivière S, Wirsing AJ, Murray DL (2006) Dietary breadth and overlap among five sympatric prairie carnivores. *Journal of Zoology*, **269**, 127-135.

Balkenhol N, Holbrook JD, Onorato D, Zager P, White C, Waits LP (2014) A multi-method approach for analyzing hierarchical genetic structures: a case study with cougars *Puma concolor*. *Ecography*, **37**, 552-563.

Barton HD, Wisely SM (2012) Phylogeography of striped skunks (*Mephitis mephitis*) in North America: Pleistocene dispersal and contemporary population structure. *Journal of Mammalogy*, **93**, 38-51.

Benjamini, Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165-1188.

Carmichael LE, Krizan J, Nagy JA, Fuglei E, Dumond M, Johnson D, Veitch A, Berteaux D, Strobeck C (2007) Historical and ecological determinants of genetic structure in arctic canids. *Molecular Ecology*, **16**, 3466-3483.

Carpenter PJ, Dawson DA, Greig C, Parham A, Cheeseman CL, Burke T (2003) Isolation of 39 polymorphic microsatellite loci and development of a fluorescently labeled marker set for the Eurasian badger (*Meles meles*) (Carnivorea: Mustelidae). *Molecular Ecology Notes,* **3**, 610-615.

Centeno-Cuadros A, Román J, Delibes M, Godoy JA (2011) Prisoners in their habitat?

Generalist dispersal by habitat specialists: a case study in southern water vole (*Arvicola sapidus*). *PloS One*, **6**, e24513.

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology,* **9**: 1657–1660.

Coppedge BR, Engle DM, Masters RE, Gregory MS (2001) Avian response to landscape change in fragmented southern Great Plains grasslands. *Ecological Applications*, **11**, 47-59.

Corander J, Siren J, Arjas E (2008) Bayesian spatial modeling of genetic population structure. *Computational Statistics*, **23**, 111–129.

COSEWIC. 2012. Canadian species at risk, May 2000. Canadian Wildlife Service, Environment Canada, Ottawa, Canada.

Courtois R, Bernatchez L, Ouellet J, Breton L (2003) Significance of caribou (*Rangifer tarandus*) ecotypes from a molecular genetics viewpoint. *Conservation Genetics*, **4**, 393-404.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959-969.

Crooks KR (2002) Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology*, **16**, 488-502.

Cullingham CI, Kyle CJ, Pond BA, White BN (2008) Genetic structure of raccoons in eastern North America based on mtDNA: implications for subspecies designation and rabies disease dynamics. *Canadian Journal of Zoology*, **86**, 947-958.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModeltest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Davis CS. Strobeck C (1998) Isolation, variability, and cross species amplification of polymorphic microsatellite loci in the family Mustelidae. *Molecular Ecology*, **7**, 1776-1778.

Domingo-Roura X, Macdonald DW, Roy MS, Marmi J, Terradas J, Woodroffe R, Burke T, Wayne RK (2003) Confirmation of low genetic diversity and multiple breeding females in a social group of Eurasian badgers from microsatellite and field data. *Molecular Ecology*, **12**, 533-539.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

Duquette JF, Gehrt SD (2014) Badger (*Taxidea taxus*) resource selection and spatial ecology in intensive agricultural landscapes. *The American Midland Naturalist*, **171**, 116-127.

Ethier DM, Laflèche A, Swanson BJ, Kocera JJ, Kyle CJ (2012) Population subdivision and peripheral isolation in American badgers (*Taxidea taxus*) and implications for conservation planning in Canada. *Canadian Journal of Zoology*, **90**, 630-639.

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

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.

Fleming, MA, Ostrander EA, Cook JA (1999) Microsatellite markers for American mink (*Mustela vison*) and ermine (*Mustela erminea*). *Molecular Ecology*, **8**, 1352–1354.

Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493-505.

Frantz AC, McDevitt AD, Pope LC et al (2014) Revisiting the phylogeography and demography of European badgers (Meles meles) based on broad sampling, multiple markers and simulations. Heredity, [in press]. doi: 10.1038/hdy.2014.45

Fritz U, Ayaz D, Buschbom J, Kami HG, Mazanaeva LF, Aloufi AA, Auer M, Rifai L, Šilić T, Hundsdörfer AK (2008) Go east: phylogeographies of *Mauremys caspica* and *M. rivulata*-discordance of morphology, mitochondrial and nuclear genomic markers and rare hybridization. *Journal of Evolutionary Biology*, **21**, 527-540.

Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L, Barnes C, Herold N, Wickham J (2011) Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, **77**, 858-864.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Garroway CJ, Bowman J, Wilson PJ (2011) Using a genetic network to parameterize a landscape resistance surface for fishers, *Martes pennanti*. *Molecular Ecology*, **20**, 3978-3988.

Godbout J, Fazekas A, Newton CH, Yeh FC, Bousquet J (2008) Glacial vicariance in the Pacific Northwest: evidence from a lodgepole pine mitochondrial DNA minisatellite for multiple genetically distinct and widely separated refugia. *Molecular Ecology*, **17**, 2463-

2475.

Goodrich JM, Buskirk SW (1998) Spacing and ecology of North American badgers (*Taxidea taxus*) in a prairie-dog (*Cynomys leucurus*) complex. *Journal of Mammalogy*, **79**, 171-179.

Graham RW, Lundelius Jr EL, Graham MA, *et al*. (1996) Spatial response of mammals to late Quaternary environmental fluctuations. *Science,* **272**, 1601-1606.

Guilday JE (1968) Grizzly bears from eastern North America. *The American Midland Naturalist*, **79**, 247-250.

Haig SM, Beever EA, Chambers SM *et al*. (2006) Taxonomic considerations in listing subspecies under the U.S. Endangered Species Act. *Conservation Biology*, **20**, 1584–1594.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Heled J, Drummnd AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.

Hengeveld R, Haeck J (1982) The distribution of abundance. I. Measurements. *Journal of Biogeography*, **9**, 303–316.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, **359**, 183-195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B (2008) The effect of inappropriate calibration: three case studies in molecular ecology. *PloS One*, **3**, e1615.

Hull JM, Strobel BN, Boal CW, Hull AC, Dykstra CR, Irish AM, Fish AM, Ernest HB (2008) Comparative phylogeography and population genetics within *Buteo lineatus* reveals evidence of distinct evolutionary lineages. Molecular Phylogenetics and *Evolution*, **49**, 988-996.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Johnson JA, Toepfer JE, Dunn PO (2003) Contrasting patterns of mitochondrial and microsatellite population structure in fragmented populations of greater prairie chickens. *Molecular Ecology*, **12**, 3335–3347.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, **4**, 782-788.

Kierepka EM, Latch EK, Swanson BJ (2012) Influence of sampling scheme on the inference of sex-biased dispersal in the American badger (*Taxidea taxus*). *Canadian Journal of Zoology,* **90**, 1231-1242.

Koblmüller S, Nord M, Wayne RK, Leonard JA (2009) Origin and status of the Great Lakes wolf. *Molecular Ecology*, **18**, 2313-2326.

Koblmüller S, Wayne RK, Leonard JA (2012) Impact of Quaternary climatic changes and interspecific competition on the demographic history of a highly mobile generalist carnivore, the coyote. *Biology Letters*, **8**, 644-647.

Kodandaramaiah U, Weingartner E, Janz N, Leski M, Slove J, Warren A, Nylin S (2012) Investigating concordance among genetic data, subspecies circumscriptions, and hostplant use in the nymphalid butterfly *Polygonia faunus*. *PloS One*, **7**, e41058.

Knaus BJ, Cronn R, Liston A, Pilgrim K, Schwartz MK (2011) Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecology*, **11**, 10.

Koen EL, Bowman J, Garroway CJ, Mills SC, Wilson PJ (2012) Landscape resistance and American marten gene flow. *Landscape Ecology*, **27**, 29-43.

Koepfli K, Deere KA, Slate GJ, Begg C, Begg K, Grassman L, Lucherini M, Veron G, Wayne RK (2008) Multigene phylogeny of the Mustelidae: resolving relationships, tempo, and biogeographic history of a mammalian adaptive radiation. *BMC Biology*, **6**, 10.

Kyle CJ, Weir RD, Newhouse NJ, Davis H, Strobeck C (2004) Genetic structure of sensitive and endangered northwestern badger populations (*Taxidea taxus taxus and T. t. jeffersonii*). *Journal of Mammalogy*, **85**, 633-639.

Lait LA, Burg TM (2013). When east meets west: population structure of a high-latitude resident species, the boreal chickadee (*Poecile hudsonicus*). *Heredity*, **111**, 321-329.

Latch EK, Rhodes Jr OE (2005) The effects of gene flow and population isolation on genetic structure of reintroduced wild turkey populations: are genetic signatures of source populations retained? *Conservation Genetics*, **6**, 981–997.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes Jr OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295-302.

Latch EK, Heffelfinger JR, Fike JA, Rhodes Jr OE (2009) Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): cryptic glacial refugia and postglacial recolonization. *Molecular Ecology*, **18**, 1730-1745.

Latch EK, Reding DM, Heffelfinger JR, Alcalá-Galván CH, Rhodes Jr OE (2014) Range-wide analysis of genetic structure in a widespread, highly mobile species (*Odocoileus hemionus*) reveals the importance of historical biogeography. *Molecular Ecology*, **23**, 3171-3190.

Long CA (1972) Taxonomic revision of the North American badger, *Taxidea taxus*. *Journal of Mammalogy,* **533**, 725-759.

Louis M, Viricel A, Lucas T, *et al*. (2014) Habitat-driven population structure of bottlenose dolpins, Tursiops truncates, in the north-east Atlantic. *Molecular Ecology*, **23**, 857-874.

Marthinsen G, Wennerberg L, Solheim R, Lifjeld JT (2009) No phylogeographic structure in the circumpolar snowy owl (*Bubo scandiacus*). *Conservation Genetics*, **10**, 923-933.

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *PNAS*, **104**, 19885-19890.

Messick JP, Hornocker MG (1981) Ecology of the badger in southwestern Idaho. *Wildlife Monographs*, **76**, 3-53.

Musiani M, Leonard JA, Cluff HD, Cormack Gates C, Mariani S, Paquet PC, Vilá C, Wayne RK (2007) Differentiation of tundra/taiga and boreal coniferous forest wolves: genetics, coat colour and association with migratory caribou. *Molecular Ecology*, **16**, 4149-4170.

Navascués M, Emerson BC (2009) Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. *Molecular Ecology*, **18**, 4390-4397.

Oksanen J, Kindt R, Legendre P, O'Hara RB (2008) vegan: community ecology package. Version 1.7–81. http://cran.r-project.org/ (accessed on 15 February 2013).

Omernik JM (1987) Ecoregions of the conterminous United States. Map (scale 1:7,500,000). *Annals of the Association of American Geographers,* **77**, 118-125.

Oyler-McCance SJ, St. John J, Knopf FL, Quinn TW (2005) Population genetic analysis of mountain plover using mitochondrial DNA sequence data. *The Condor*, **107**, 353-362.

Pesole G, Gissi C, De Chirico A, Saccone C (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution*, **48**, 427-434.

Placyk Jr JS, Burghardt GM, Small RL, King RB, Casper GS, Robinson JW (2007) Post-glacial recolonization of the Great Lakes region by the common garter snake (*Thamnophis sirtalis*) inferred from mtDNA sequences. *Molecular Phylogenetics and Evolution*, **43**, 452-467.

Posada D, Crandall KA (2001) Intraspecific gene genealogies:trees grafting into networks. *Trends in Ecology and Evolution*,**16**, 37–45.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155**, 945–959.

Pritchard JK, Wen W (2003) Documentation for STRUCTURE software: version 2. Available from http://pritch.bsd.uchicago.edu (accessed 12 May 2012).

Pulagrín-Restrepo PC, Burg TM (2012) Genetic signals of demographic expansion in downy woodpecker (Picoides pubescens) after the last North American glacial maximum. *PloS One*, **7**, e40412.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092-2100.

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

Reding DM, Bronikowski AM, Johnson WE, Clark WE (2012) Pleistocene and ecological effects on continental-scale differentiation in the bobcat (*Lynx rufus*). *Molecular Ecology*, **21**, 3078-3093.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539-542.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58-62.

Row JR, Gomez C, Koen EL, Bowman J, Murray DL, Wilson PJ (2012) Dispersal promotes high gene flow among Canada lynx populations across mainland North America. *Conservation Genetics*, **13**, 1259-1268.

Rowe KC, Heske EJ, Brown PW, Paige KN (2004) Surviving the ice: northern refugia and postglacial colonization. *PNAS*, **101**, 10355-10359.

Runck AM, Cook JA (2005) Postglacial expansion of the southern red-backed vole (*Clethrionomys gapperi*) in North America. *Molecular Ecology*, **14**, 1445-1456.

Sabatino SJ, Routman EJ (2009) Phylogeography and conservation genetics of the hellbender salamander (Cryptobranchus alleganiensis). *Conservation Genetics*, **10,** 1235-1246.

Sacks BN, Mitchell BR, Williams CL, Ernest HB (2005) Coyote movements and social structure along a cryptic population genetic subdivision. *Molecular Ecology*, **14**, 1241-1249.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. **Genetics**, *152*, 1079–1089.

Schwartz MK, Copeland JP, Anderson NJ, Squires JR, Inman RM, McKelvey KS, Pilgrim KL, Waits LP, Cushman SA (2009) Wolverine gene flow across a narrow climatic niche. *Ecology*, **90**, 3222-3232.

Shafer ABA, Cullingham CI, Coté SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**, 4589-4621.

Shafer ABA, Coté SD, Coltman DW (2011) Hot spots of genetic diversity descended from multiple Pleistocene refugia in an alpine ungulate. *Evolution*, **65**, 125-138.

Shapiro B, Ho SYW (2014) Ancient hyaenas highlgith the old problem of estimating evolutionary rates. *Molecular Ecology*, **23**, 499-501.

Short Bull RA, Cushman SA, Mace R, Chilton T, Kendall KC, Landguth EL, Schwartz MK, McKelvey K, Allendorf FW, Luikart G (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology*, **20**, 1092-1107.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mtDNA sequences in stable and

exponentially growing populations. *Genetics*, **129**, 555–562.

Smith BT, Escalante P, Hernández Baños BE, Navarro-Sigüenza AG, Rohwer S, Klicka J (2011) The role of historical and contemporary processes on phylogeographic structure and genetic diversity in the northern cardinal, *Cardinalis cardinalis*. *BMC Evolutionary Biology*, **11**, 136.

Sovada MA, Roaldson JM, Sargeant AB (1999) Foods of American badgers in west-central Minnesota and southeastern North Dakota during the duck nesting season. *The American Midland Naturalist*, **142**, 410-414.

Steffan-Dewenter I, Tscharntke T (2002) Insect communities and biotic interactions on fragmented calcareous grasslands-a mini review. *Biological Conservation*, **104**, 275-284.
Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist*, **166**, 581-591.

Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society: Biological Sciences*, **227**, 661-671.

Swofford DL (2003) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b10a. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Tammeleht E, Remm J, Korsten M, Davidson J, Tumanov I, Saveljev A, Männil P, Kojola I, Saarma U (2010) Genetic structure in large, continuous mammal populations: the example of brown bears in northwestern Europe. *Molecular Ecology,* **19**, 5359–5370.

Tavaré S (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. Pp. 57-86 in R. M. Miura, ed. Some mathematical questions in biology - DNA sequence analysis. American Mathematical Society, Providence, RI.

Taylor ZS, Hoffman SMG (2010) Mitochondrial DNA genetic structure transcends natural boundaries in Great Lakes populations of woodland deer mice (*Peromyscus maniculatus gracilis*). *Canadian Journal of Zoology*, **88**, 404-415.

Teacher AGF, Thomas JA, Barnes I (2011) Modern and ancient red fox (Vulpes vulpes) in Europe show an unusual lack of geographical and temporal structuring, and differing responses within the carnivores to historical climatic change. *BMC Evolutionary Biology*, **11**, 214.

Trumbo DR, Spear SF, Baumsteiger J, Storfer A (2013) Rangewide landscape genetics of an endemic Pacific northwestern salamander. *Molecular Ecology*, **22**, 1250-1266.

van Els P, Spellman GM, Tilston Smith B, Klicka J (2014) Extensive gene flow characterizes the phylogeography of a North American migrant bird: black-headed grosbeak (*Pheucticus melanocephalus*). *Molecular Phylogenetics and Evolution*, **78**, 148-159.

Vilà C, Amorim IR, Leonard JA, Posada D, Castroviejo J, Petrucci-Fonseca F, Crandall KA, Ellegren H, Wayne RK (1999) Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus*. *Molecular Ecology*, **8**, 2089-2103.

vonHoldt BM, Pollinger JP, Earl DA, *et al*. (2011) A genome-wide perspective on the evolutionary history enigmatic wolf-like canids. *Genome Research*, **21**, 1294-1305.

Vucetich JA, Waite TA (2003) Spatial patterns of demography and genetic processes across the species' range: null hypotheses for landscape conservation genetics. *Conservation Genetics*, **4**, 639-645.
Warner, R.E., and Ver Steeg. B. 1995. Illinois badger studies. Division of Wildlife Resources, Illinois Department of Natural Resources, Springfield, Illinois.

Weckworth BV, Musiani M, DeCesare NJ, McDevitt AD, Hebblewhite M, Mariani S (2013) Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. *Proceedings of the Royal Society: Biological Sciences*, **280**, 20131756.

Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society: Biological Sciences*, **271**, 1881-1887.

Whitlock C, Sarna-Wojcicki AM, Bartlein PJ, Nickmann RJ (2000) Environmental history and tephrostratigraphy at Carp Lake, southwestern Columbia Basin, Washington, USA.

Wisely SM, Statham MJ, Fleischer RC (2008) Pleistocene refugia and Holocene expansion of a grassland-dependent species, the black-footed ferret (*Mustela nigripes*). *Journal of Mammalogy*, **80**, 87-96.

Wooding S, Ward R (1997) Phylogeography and Pleistocene evolution in the North American black bear. *Molecular Biology and Evolution*, **14**, 1096-1105.

Wright S (1931) Evolution in Mendelian population. *Genetics*, **16**, 97-159.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114-138.

Zamudio KR, Savage WK (2003) Historical isolation, range expansion, and secondary contact of two highly divergent mitochondrial lineages in spotted salamanders (*Ambystoma maculatum*). *Evolution*, **57**, 1631-1652.

Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society: Biological Sciences,* **271**, 561-564.

CURRICULUM VITAE

Education

B.S., Michigan State University-December 2006
Major: Zoology, Honors

M.S., Central Michigan University-May 2009
Major: Biology, Conservation Biology

Dissertation Title: Landscape genetics of the American badger: understanding challenges with elusive species