

ORIGINAL ARTICLE

# Fine-scale landscape genetics of the American badger (*Taxidea taxus*): disentangling landscape effects and sampling artifacts in a poorly understood species

EM Kierepka and EK Latch

Landscape genetics is a powerful tool for conservation because it identifies landscape features that are important for maintaining genetic connectivity between populations within heterogeneous landscapes. However, using landscape genetics in poorly understood species presents a number of challenges, namely, limited life history information for the focal population and spatially biased sampling. Both obstacles can reduce power in statistics, particularly in individual-based studies. In this study, we genotyped 233 American badgers in Wisconsin at 12 microsatellite loci to identify alternative statistical approaches that can be applied to poorly understood species in an individual-based framework. Badgers are protected in Wisconsin owing to an overall lack in life history information, so our study utilized partial redundancy analysis (RDA) and spatially lagged regressions to quantify how three landscape factors (Wisconsin River, Ecoregions and land cover) impacted gene flow. We also performed simulations to quantify errors created by spatially biased sampling. Statistical analyses first found that geographic distance was an important influence on gene flow, mainly driven by fine-scale positive spatial autocorrelations. After controlling for geographic distance, both RDA and regressions found that Wisconsin River and Agriculture were correlated with genetic differentiation. However, only Agriculture had an acceptable type I error rate (3–5%) to be considered biologically relevant. Collectively, this study highlights the benefits of combining robust statistics and error assessment via simulations and provides a method for hypothesis testing in individual-based landscape genetics.

*Heredity* (2016) **116**, 33–43; doi:10.1038/hdy.2015.67; published online 5 August 2015

## INTRODUCTION

Landscape genetics has become an increasingly important tool for conservation and management by identifying landscape factors that influence genetic connectivity across heterogeneous landscapes. Loss of genetic connectivity across landscapes can depress genetic diversity and potentially increase extinction risks (Epps *et al.*, 2005; Dixo *et al.*, 2009; Clark *et al.*, 2010; Ernest *et al.*, 2014), so landscape genetics can provide information to help understand and potentially mitigate the effects of land cover change in natural populations (for example, fragmentation, habitat loss, anthropogenic disturbance; Segelbacher *et al.*, 2010). To understand how landscape heterogeneity may impact genetic connectivity within a focal species, researchers typically utilize field studies (for example, habitat selection or occupancy) to develop landscape resistance models. 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 and Waits, 2010; Garroway *et al.*, 2011), climatic conditions (Row *et al.*, 2014) and anthropogenic barriers (Blanchong *et al.*, 2008; Latch *et al.*, 2011) also strongly influence gene flow. When field data corroborates correlations from landscape genetics (for example, 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 (for example, Cushman *et al.*, 2006; Schwartz *et al.*, 2009; Shafer *et al.*, 2012), application to poorly understood species can be difficult. Ideally, investigators should develop specific hypotheses about how landscape heterogeneity impacts gene flow in a focal species within the study area and parameterize resistance models accordingly (for example, Balkenhol *et al.*, 2009; Anderson *et al.*, 2010). For many 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) despite criticism over decreased accuracy (Pearce *et al.*, 2001; Clevenger *et al.*, 2002), inherent biases and inability to assess accuracy in expert opinion (Spear *et al.*, 2010; Zeller *et al.*, 2012). One potential solution to relying on expert opinion in poorly understood species is to employ alternative statistical methods that do not require extensive parameterization of landscape resistance models.

Ordination and regression methods (for example, Legendre and Legendre, 1998; Fortin and Legendre, 2010; Wang, 2013) have two main advantages that could alleviate problems associated with a lack of relevant field data. First, these methods offer considerable flexibility in the type of landscape and genetic variables that can be evaluated, negating reliance on *a priori* parameterization of landscape resistance

hypotheses. Furthermore, ordination and regression techniques are considered more robust than correlation statistics (for example, Mantel tests; Fortin and Legendre, 2010), and therefore, may limit type I or II errors (Balkenhol *et al.*, 2009; Kierepka and Latch, 2015). Recent studies have demonstrated the utility of ordination 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 (that is, between genetic clusters, Reding *et al.*, 2012; or between sampled areas, Blanchong *et al.*, 2008; Robinson *et al.*, 2012). Extending such methods to poorly understood species within a single study area or population requires individual-based approaches that can simultaneously test biologically relevant patterns in gene flow as well as disentangle landscape effects from spurious statistical correlations.

All individual-based landscape genetic statistics, even ordination and regression techniques, are likely to suffer from some degree of error owing to spatially biased sampling (that is, sampling individuals in accessible areas or opportunistically; Storfer *et al.*, 2007). Multiple statistics in landscape genetics already have elevated type I error rates (that is, false significance; Balkenhol *et al.*, 2009; Graves *et al.*, 2013; Guillot and Rousset, 2013), and with spatially biased sampling further creating non-random genetic variation (Schwartz and McKelvey, 2009; Oyler-McCance *et al.*, 2013), authors need a method to prevent erroneous conclusions about gene flow. Gene flow simulations provide a means of replication within a single landscape and control over processes that result in observed genetic variation (for example, Epperson *et al.*, 2010; Landguth *et al.*, 2010) and thus can quantify how often statistics falsely identify focal landscape factors as significant (that is, type I error rates for each landscape factor). By quantifying type I errors using gene flow simulations, investigators can better understand how landscape heterogeneity impacts gene flow in poorly understood species and separate those effects from sampling artifacts.

The American badger (*Taxidea taxus*) is one of the most poorly understood mesocarnivore species in North America. Badgers inhabit treeless habitats across much of central and western North America, but much of their life history is unknown due to their nocturnal, semifossorial life style. Field investigations have occurred in badgers (for example, British Columbia, Apps *et al.*, 2002; Ohio, Duquette and Gehrt 2014; Wyoming, Messick and Hornocker 1981), but few commonalities exist owing to high variability in ecological traits, such as abundance, territoriality and habitat selection. This variability is particularly concerning because differences in life history also create disparate patterns in gene flow across landscapes and spatial scales (Short Bull *et al.*, 2011; Zeller *et al.*, 2012) making development of landscape resistance models in understudied species difficult. In this study, we focused on badgers in Wisconsin, an area where population characteristics such as abundance, habitat preferences and movement behavior are unknown. Combined with the unavoidable spatially biased sampling, the overall lack of relevant life history data on badgers within Wisconsin presents a substantial challenge for deciphering between potential errors (for example, improper parameterization and type I errors) and actual patterns in gene flow.

In this study, we genotyped 233 individual badgers at 12 microsatellite loci to examine the utility of integrating ordination and regression techniques with simulations for overcoming challenges associated with limited life history and spatially biased sampling. To identify landscape factors that influence badger gene flow, this study used ordination and regression techniques to test whether 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. Ordination

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 required a statewide citizen-based effort where citizens reported badger sightings such as active burrows and road-kills (Kierepka, 2014). This method of sampling resulted in samples clumped around populated areas, a factor that could lead to erroneous conclusions about gene flow (Schwartz and McKelvey, 2009; Oyler-McCance *et al.*, 2013). Therefore, we also performed gene flow simulations to calculate how often our chosen statistics falsely identified landscape effects (that is, type I errors) to separate significant results created by sampling and actual landscape effects.

## MATERIALS AND METHODS

### Study area

Our study area (110 745 km<sup>2</sup>) encompasses the state of Wisconsin in the Upper Midwest, USA. 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 to 2014 (Kierepka, 2014) and Wisconsin Department of Natural Resources mammal surveys from 1987 to 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 owing to the Mississippi River and Great Lakes (Kierepka, 2014).

### Sample collection

As badgers are protected from harvest in Wisconsin as a Species with Information Needs (Wisconsin Department of Natural Resources, 2008), hair ( $n = 111$ ) and tissue ( $n = 139$ ) samples ( $n = 250$  total) were collected from 2004 to 2013. All badgers sampled before 2009 were deceased animals, from either road-kills or incidental captures ( $n = 47$ ), whereas samples collected from 2009 to 2013 were a mix of live animals ( $n = 106$ ) and road-kills ( $n = 90$ ). Live animals consisted of either live captures ( $n = 25$ ) or via hairs collected from active burrows ( $n = 81$ ).

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 (that is, activity within burrow reported within 24 h) and waiting overnight for the animal to pass under the snare. A successful hair collection typically contained 20–50 long banded hairs with intact roots, ensuring a sufficient number of hairs for molecular analysis. The majority of hair snares (78/81) were deployed during March–August, when family groups were together based on the citizen monitoring program in Wisconsin. Road-kills were collected from March to October, but peaks in road-kills occurred from August to October, roughly corresponding to the hypothesized breeding season and dispersal of juveniles.

### Laboratory methods

DNA was extracted using Qiagen DNEasy Blood and Tissue Kits (QIAGEN Inc., Valencia, CA, USA) for both tissues and hairs. Extractions used a 1-mm<sup>3</sup> piece of tissue or 10–20 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 (Supplementary Table S1) developed in American badger (Tt-1, Tt-2, Tt-3 and Tt-4; Davis and Strobeck, 1998), American mink (*Neovison vison*; Mvis072; Fleming *et al.*, 2002), American marten (*Martes americana*; Ma-1; Davis and Strobeck, 1998) and European badger (*Meles meles*; Mel112, Mel101, Mel111, Mel108, Mel14 and Mel1; Carpenter *et al.*, 2003; Domingo-Roura *et al.*, 2003). Multiplex PCRs were conducted in sets of 2–4 primers in 10- $\mu$ l reaction volumes (Supplementary Table S1). Amplified products were genotyped on an ABI 3730 DNA Analyzer (Life Technologies, Grand Island, NY, USA) 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,  $n = 17$ ) 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 six instances of allelic dropout (all hairs) in the 918 repeated genotypes (0.65% error rate). Any individuals with  $< 10$  genotypes were then culled from the data set. As a final step to remove highly related individuals, we calculated relatedness values ( $r$ ) between all individuals in the program SPAGED1 v. 1.2 (Hardy and Vekemans, 2002). Any  $r$  that was outside three times the interquartile range of all  $r$  values (that is, outlier in distribution) was considered highly related ( $r > 0.645$ ). One individual from each highly related pair was then removed from the data set ( $n = 3$  individuals removed), resulting in 233 individuals in our total data set (17 samples culled; Supplementary Table S2; Figure 1).

### Population structure

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 10 independent runs (100 000 Markov chain Monte Carlo (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  $\Delta K$  (Evanno *et al.*, 2005) because likelihood values plateaued and variances among runs grew larger at values of  $K$  above the optimum (Pritchard *et al.*, 2000; Supplementary

Figure S1a). Once the optimal  $K$  was identified, 10 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 10 runs were calculated in CLUMPP (Jakobsson and 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  $K = 1-10$  (10 replicates per  $K$ ) using the 'Spatial Clustering of Individuals' option in BAPS. Location information for each individual was recorded in latitude and longitude coordinates or using the 'Create Random Points' function within ArcMap v. 10.1 for individuals with less precise location information (that is, counties, Public Land Survey System locations;  $n = 42$ ). 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 diveRcity (Keenan, 2013; R Core Team, 2013). Deviations from Hardy-Weinberg and linkage equilibria were calculated in GENEPOP (Raymond and Rousset, 1995) using a corrected alpha for multiple tests ( $\alpha = 0.012$ ; false discovery rate; Benjamini and Yekutieli, 2001), and null allele frequencies were calculated in MICRO-CHECKER (van Oosterhout *et al.*, 2004).

### Barriers to gene flow

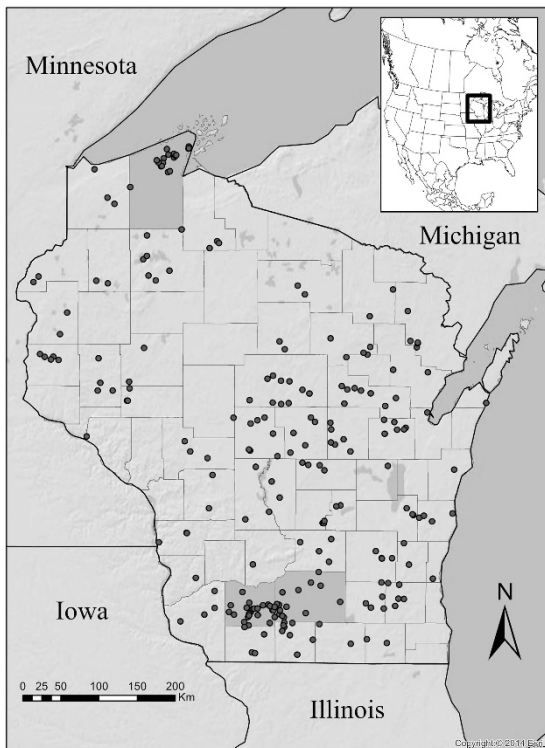
To test for the effect of geographic distance on genetic differentiation, we quantified patterns of IBD within Wisconsin using two complementary approaches. We used a simple Mantel test to test for an association between matrices of pairwise genetic and geographic distances and distance-based redundancy analysis (dbRDA) to test for significant effects of geography (latitude and longitude) on the distribution of genetic variation across the study area. Both statistical tests require pairwise genetic distances, so we calculated genetic distances between all individuals (Rousset's  $a$ ; Rousset, 2000) in the program SPAGED1. All IBD tests (simple Mantel test and dbRDA) 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 (that is, panmixia) within 5-km distance categories. Individuals separated by small geographic distances are expected to exhibit positive spatial autocorrelations (that is, 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 biases (Figure 1).

Population structure is often influenced by discrete barriers (isolation-by-barrier) in addition to geographic distance (IBD). We used a partial Mantel test to determine correlational significance between a genetic distance matrix (Rousset's  $a$ ) 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.

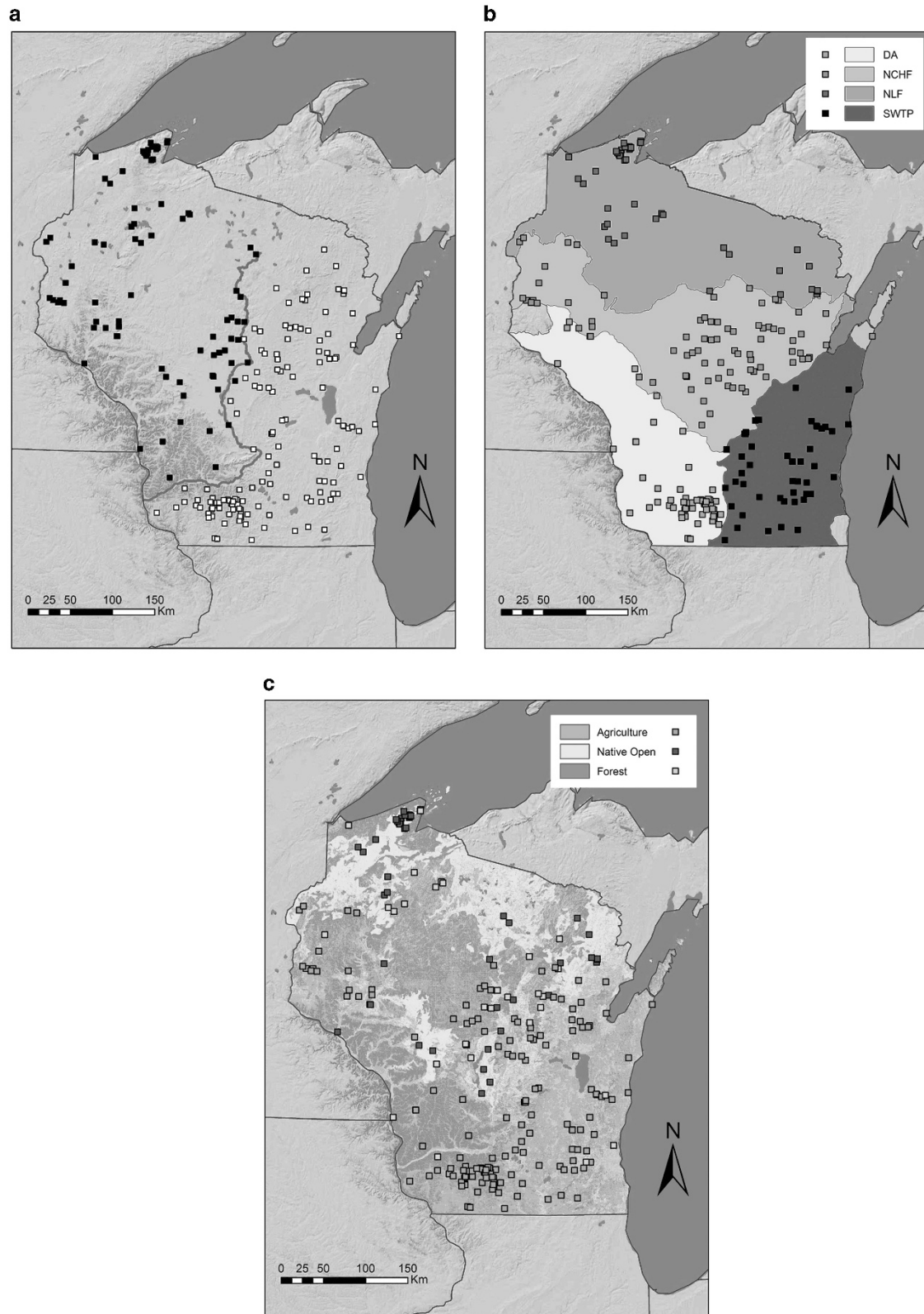
### Landscape factor derivation

In nature, many different landscape factors can work in tandem to create observed patterns of genetic variation (for example, 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). We used level III Ecoregions (Figure 2) to define broad-scale regions of similar abiotic (that is, climate and soil) and biotic assemblages (in Wisconsin: Northern Lakes and Forest (NLF); North Central Hardwood Forest (NCHF), Driftless Area (DA),



**Figure 1** Sampling locations of 233 badgers with sufficient genetic data for analysis. Counties within Wisconsin are outlined in gray and the three most heavily sampled counties (Bayfield, Dane and Iowa counties) are labeled and highlighted in gray. All states surrounding Wisconsin are labeled.





**Figure 2** Predictor variables for each georeferenced badger describe its location according to three landscape features (Wisconsin River, level III Ecoregion and land cover; gray in print). 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 the percentage of land cover within a circular buffer surrounding each badger's location (most predominant land cover class given by point's color; c). A full color version of this figure is available at the *Heredity* journal online.

and Southeastern Wisconsin Till Plain (SWTP); Omernik, 1987). The full description of how level III Ecoregions are defined is available in (Omernik, 1987), but briefly, two ecoregions (NLF and NCHF) are largely forested, whereas DA and SWTP are dominated by agriculture. NLF has the least agriculture among the included ecoregions with more sandy plains and coniferous forests than NCHF. NCHF, in contrast, contains a mosaic of forest, wetland and agricultural areas with varied soils types. SWTP is primarily agriculture with silt-loam soils. DA was left unglaciated, resulting in greater variability in topography (prairies at higher elevations and forests within valleys) and soil types (exposed bedrock to clay).

In addition to level III Ecoregions, 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 (Figure 2). 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.

We also attempted to include anthropogenic infrastructure as variables, including roads (proportion of roads or road densities in buffers) and urban areas (proportion roads and urban combined or urban alone), but little variability was observed in these values (175–203/233 observations had  $< 0.100$  urban, mean =  $0.0733$ – $0.0879$  across buffer widths). Furthermore, unlike many species that exhibit genetic differentiation according to roads (for example, Cushman and Lewis, 2010; Frantz *et al.*, 2010a, b; Frantz *et al.*, 2012; Galpern *et al.*, 2012; Robinson *et al.*, 2012), American badgers do not prefer habitats with ample cover (that is, forests) or avoid even large roads (Apps *et al.*, 2002). Badgers are thought to prefer roadside habitats because of the loose soils and abundant rodent prey (Messick and Hornocker, 1981; Apps *et al.* 2002). Although many of our samples were road-kills, a potential mechanism for a barrier effect, burrows in Wisconsin were often observed on roadsides and photo evidence of badger activity included several instances of individuals crossing large highways. The low variability of roaded habitats within each buffer and no apparent mechanisms for barrier effects precluded the incorporation of roads within our landscape genetic analysis.

### Landscape genetic analysis

Ordination 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 1000 permutations. We used an inverse distance-weighting network so that all badgers were considered neighbors. The first two sPCA axes explained the most variation (eigenvalues =  $0.0889$  and  $0.0571$ , all others  $< 0.0459$ ; Supplementary Figure S3), so the spatially lagged scores for these axes were retained as dependent variables in landscape genetic analysis.

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 RDA, a constrained ordination technique that is the multivariate analog to simple linear regression (Legendre and Legendre, 1998). Previous simulation studies in both population- (Fortin and Legendre, 2010) and individual-based (Kierepka and Latch, 2015) studies have demonstrated that RDA has much greater power than Mantel tests to detect relationships in autocorrelated data. In individual-based studies, however, RDA can suffer from type I errors (that is, false significance; Kierepka and Latch, 2015), so additional steps are needed to control for any potential errors. The spatially lagged scores from the two retained sPCA axes were analyzed together as explanatory variables with both latitude and longitude in the conditional matrix (that is, variables to be controlled for in final models). The barrier (Wisconsin River) and ecological (Ecoregion or Agriculture) variables were included as predictors (See Figure 2 for classifications). To prevent correlations in our explanatory variables, we calculated variance inflation factors (VIF) for our predictor variables using the function 'vif.cca', but no such correlations were found (all  $VIF < 1.000$ ). VEGAN provides a stepwise model selection

procedure to identify the variables that best explained genetic differences among individuals in the function 'ordstep' followed by significance testing in 'anova.cca' (10 000 permutations for both functions; 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 the spatially lagged score of each sPCA axis individually, which provides further evidence for specific influences on patterns found within the sPCA. It is unknown how spatially lagged regression performs in individual-based landscape genetics, because the limited studies have utilized this technique in a population-based framework (for example, Robinson *et al.*, 2012). Spatially lagged regression models control for IBD by including a spatial weighting matrix ( $W_{ij}$ ) and  $\rho$ , a parameter that accounts for the lack of independence between individuals (Legendre and 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:  $G_i = \rho \times W_{ij} \times G_j + \beta \times X + \epsilon$  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 ( $\beta$ ), and residual error ( $\epsilon$ ; Legendre and Legendre, 1998). A LaGrange Multiplier test then tests for spatial autocorrelation within residuals of the regressions where a significant test indicates autocorrelation remains within the data (Bivand *et al.*, 2011). One major advantage of spatial regression is the ability to conduct model selection procedures such as Akaike Information Criterion (AIC) while providing regression coefficients and measures of variability that are largely unavailable in Mantel tests (Fortin and Legendre, 2010; Guillot and Rousset, 2013). Model selection followed the method by Burnham and Anderson's (2002) where models with a  $\Delta AIC < 2.0$  were considered candidate models for explaining an sPCA axis. Parameter estimates were produced through model averaging of all candidate models where significant parameters do not include zero (Burnham and Anderson, 2002). All spatial regression methods were performed using the 'lagsarlm' function in R package SPDEF (Bivand *et al.*, 2011).

### Type I error assessment

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 and Cushman, 2009). CDPOP allows manipulation of both landscape resistance to gene flow and assignment of demographic characteristics. We aimed to create a single population that exhibited IBD, so the underlying landscape was homogeneous. Both sexes had a similar dispersal regime because strong sex-biased dispersal has not been recorded in badgers (Messick and Hornocker, 1981; Hoodicoff, 2003; Kierepka *et al.*, 2012). Both sexes dispersed according to an inverse square distribution, because this distribution ensures short distance dispersal is much more common than long distance. Mating was sexual with replacement (that is, both sexes could breed multiple times) and generations were overlapping.

The actual census size and population density of badgers within Wisconsin is unknown, so the majority of individuals within simulated populations were placed randomly on Wisconsin's landscape excluding urban and open water habitats. The initial population for all simulations consisted of 1400 individuals, 1167 with random locations and 233 with geographic coordinates identical to our empirical data set. After 250 generations, we sampled the 233 individuals that had the same geographic locations as the empirical data set. Therefore, we had 100 independent populations that were sampled identically to our empirical data set that only exhibited IBD (verified via simple Mantel tests; all  $r > 0.0512$ , all  $P < 0.025$ ).

For each simulated population, we then performed the same sPCA, partial RDA and spatially lagged regression analyses as the empirical data set. A type I error occurred when a barrier (Wisconsin River) or ecological (Ecoregion, land cover) factor was falsely significant as in  $> 5$  of the 100 simulated population's ( $\alpha = 0.05$ ).

## RESULTS

### Population structure

In the BAPS analysis, the optimal solution was  $K=1$  (−11197.19). The STRUCTURE analysis indicated stronger support for  $K=2$  than for  $K=1$  and a second small peak at  $K=4$  (Supplementary Figure S1a). A second peak in  $\Delta K$  can sometimes be indicative of additional substructure (as observed in hierarchical island models; Evanno *et al.*, 2005), but iterative runs of the  $K=2$  scheme failed to find any additional substructure (that is, all  $q$ -values ranged from 0.4 to 0.6). Furthermore, an inspection of the STRUCTURE assignments at  $K=2$  and  $K=4$  revealed gradient patterns instead of any discrete barriers, and relatively equal probabilities of assignment to each cluster (Supplementary Figure S1b). As a final check, we evaluated  $F_{ST}$  between the two putative clusters from STRUCTURE, and it was not significant ( $F_{ST}=0.008$ ,  $P=0.433$ ). The resultant gradient in  $q$ -values at  $K=2$ , insignificant  $F_{ST}$  and results from BAPS suggests that the inferred structure was likely an artifact of spatial autocorrelation (Schwartz and McKelvey, 2009) or IBD (Frantz *et al.*, 2009).

We observed a heterozygote deficiency within Wisconsin over all loci in a global analysis (Table 1) and for seven individual loci (all  $P<0.001$ ). These deviations from Hardy–Weinberg equilibrium are expected if any deviation from panmixia exists within the data set or if there are genotyping errors. All loci were highly polymorphic ranging from 7 to 15 alleles per locus (average=11.58 alleles/locus) and showed no linkage disequilibrium (all  $P>0.025$ ). Although MICRO-CHECKER identified an overall excess of homozygotes, the calculated potential null allele frequencies were low (0.000–0.1548) and distributed across seven loci. It is therefore likely that the homozygote excess we observed is due to a deviation from panmixia rather than genotyping error, and thus we retained all alleles in subsequent analyses to maximize explanatory power (Dharmarajan *et al.*, 2013). Although it is improbable that low frequency null alleles were present in our data set, it is unlikely that they would bias our conclusions of this study given their low frequency and the specific goals of this study (Dakin and Avise, 2004; Chapuis and Estoup, 2007; Carlsson, 2008).

### Barriers to gene flow

Results from the dbrDA analysis supported the Bayesian clustering and Hardy–Weinberg equilibrium analyses that indicated a role for geography driving gene flow patterns. However, the dbrDA analysis

**Table 1** Locus-specific summary of genetic variation for  $n=233$  badgers in Wisconsin

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

Abbreviation: CI, confidence interval. 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 Hardy–Weinberg equilibrium at a corrected  $\alpha=0.012$ .

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 ( $r=0.012$ ,  $P=0.312$ ; Supplementary Figure S2). 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 ( $\leq 25$  km) were smaller than expected under panmixia in the spatial autocorrelation (Supplementary Figure S3). 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 53/90 pairwise comparisons (Dane=3, Iowa=21 and Bayfield=29) under 5 km occurred. However, most of these samples within these comparisons ( $n=37$  individuals) were road-killed individuals ( $n=10$ ) sampled across years or live captured adult animals at separate burrows ( $n=11$ ). The remaining samples were hair snared during May 2010 in Bayfield County ( $n=16$ ), a time where badger kits are still dependent on their mothers. Based on our protocol of only utilizing long hairs typical of adults in extractions, it is unlikely we sampled mothers and their kits at separate burrows. As a final test, 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 isolation-by-barrier 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 ( $r=-0.0266$ ,  $P=0.992$ ).

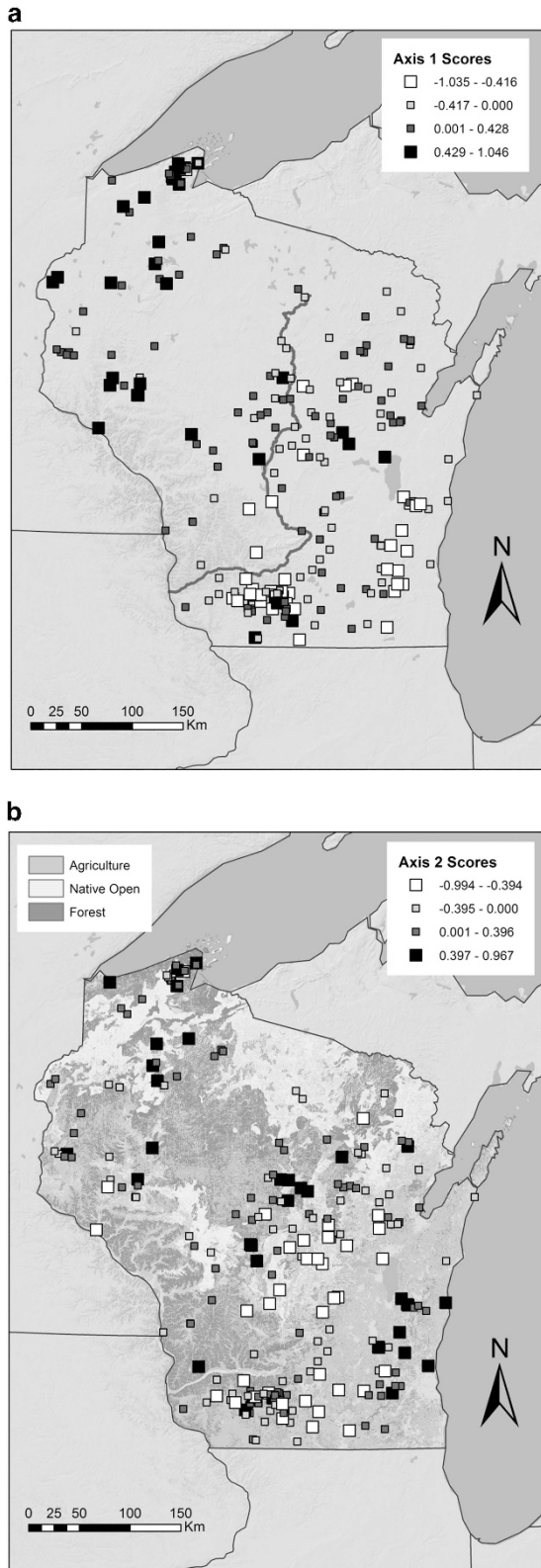
### Landscape genetic analysis

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 3). 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; Supplementary Figure S4). 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 =  $-0.225 \pm 0.144$ ; Table 2). However, observed models also had relatively high residual spatial autocorrelations (LaGrange Multiplier test: = 1.967–4.768,





**Figure 3** Spatially lagged scores for the first two sPCA axes for Wisconsin badgers. Scores from axes 1 (a) and 2 (b) were correlated with the Wisconsin River and the percentage of Agriculture (Ag), respectively. Dark colors (black and dark gray) represent negative sPCA scores while positive values are light in color (white and light gray). More extreme values in sPCA axes are displayed with larger squares.

**Table 2** Model selection results for spatially lagged regressions on sPCA axes 1 and 2

Model	AIC	$\Delta AIC$	$w_i$
<i>Axis 1</i>			
<b>Eco+River</b>	<b>169.66</b>	<b>0.00</b>	<b>0.499</b>
<b>Ag+Eco+River</b>	<b>171.46</b>	<b>1.80</b>	<b>0.203</b>
<b>River</b>	<b>171.49</b>	<b>1.83</b>	<b>0.200</b>
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
<i>Axis 2</i>			
<b>Ag</b>	<b>216.68</b>	<b>0.00</b>	<b>0.482</b>
<b>Ag+River</b>	<b>218.20</b>	<b>1.52</b>	<b>0.225</b>
<b>Ag+Eco</b>	<b>218.52</b>	<b>1.84</b>	<b>0.192</b>
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

Abbreviation: AIC, Akaike Information Criterion. Top models ( $\Delta AIC < 2.0$ ; 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.

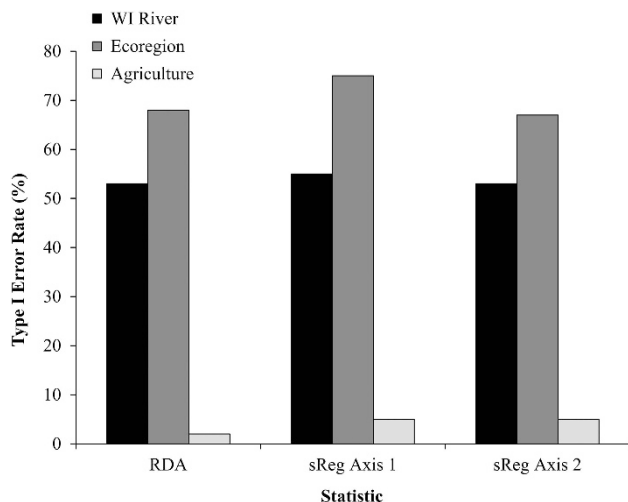
$P = 0.0290-0.116$ ), suggesting that IBD was also associated with Axis 1. All top models for Axis 2 included Agriculture (Agriculture, Agriculture + River, Agriculture + Ecoregion; Table 2). Model averaging between these three models indicated that Agriculture was the only significant variable (model averaged parameter:  $-0.318 \pm 0.284$ ). 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.

#### Type I error assessment

Simulations revealed similar type I error rates for partial RDA and spatially lagged regression (Figure 4). 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 cutoff. A likely cause of high error rates was that residual autocorrelation (that is, 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 data set 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  $< 5\%$  of populations regardless of the test (3–5/100 populations; Figure 4). 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 data set and represents an ecological variable that likely impacts gene flow.

#### DISCUSSION

There are many challenges associated with poorly understood species that can affect conclusions about how features of the landscape influence gene flow. In badgers, both limited life history data and



**Figure 4** Percentage 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.

spatially biased sampling were obstacles that could potentially confound our interpretation of landscape genetic statistics. Our study demonstrates that combining ordination 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 calculation of connectivity indices (Kierepka and Latch, 2015), 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 data set indicated that geographic distance is the strongest influence on badger gene flow within Wisconsin, with Agriculture having a lesser role.

#### Landscape genetics of badgers in Wisconsin

Geographic distance was the primary influence on gene flow in badgers, as evidenced by dBRDA, spatial autocorrelations and sPCA. 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 data set, a likely result of 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 and Hornocker, 1981; Kierepka *et al.*, 2012). Dispersal regimes can vary according to habitat quality where dispersal is more restricted in suitable habitat (for example, 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 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 through 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 (for example, Cegelski *et al.*, 2006; Schwartz *et al.*, 2009; 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 analyses 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 (that is, 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 data set 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 and Hornocker, 1981; Warner and Ver Steeg, 1995; Duquette and 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 except in the northwestern open habitats. In southern Wisconsin, Native Open habitats were often interspersed with Agriculture, but spatial data may have lacked the resolution to capture fine-scale habitats necessary for dispersal (Anderson *et al.*, 2010) similar to native habitat along fencerows (Duquette and Gehrt, 2014). Alternatively, dispersal through sub-optimal habitats via density-dependent dispersal (for example, *Martes pennanti*; Carr *et al.*, 2007) or compensatory movements (Rosenberg *et al.*, 1998) could also explain why Agriculture enhanced gene flow. For badgers, unavailable Native Open habitats (Carr *et al.*, 2007) and/or faster movements through Agriculture during dispersal (for example, Schultz and Crone, 2001; Dickson *et al.*, 2005; Rizkalla and Swihart, 2007) could produce high genetic similarity among individuals within agricultural habitats and surrounding suitable habitats (for example, Native Open or pasturelands in central and southwestern Wisconsin, respectively). Under this scenario, we would also expect 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 without corroborating field data, but regardless of the mechanism, badgers appear able to successfully disperse through sub-optimal habitats.

American badger's tolerance for fragmentation and overall high gene flow differs greatly from the more well-studied European badger (*Meles meles*) populations in the United Kingdom and Ireland. In European badgers, multiple authors have recorded evidence for restricted dispersal owing to both natural (that is, rivers; Sleeman *et al.*, 2009; Frantz *et al.*, 2010b) and anthropogenic (roads; Clark *et al.*, 1998; Frantz *et al.*, 2010a) barriers, as well as strong fine-scale IBD created by philopatry within dense populations (Pope *et al.*, 2006; Frantz *et al.*, 2010a). American badgers, in contrast, do not avoid large roads (Apps *et al.*, 2002), do not form social groups and occur at lower densities than UK badger populations, so the difference in gene flow patterns likely stems from their disparate ecologies.

#### Utility of ordination and regressions in landscape genetics

When interpreted properly through rigorous error testing, ordination and regression statistics are well suited to poorly understood species such as badgers in individual-based landscape genetics. One benefit of ordination 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 ordination and regression techniques (for example, 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 ordination and regression techniques are equally useful in individual-based studies with appropriate choices in landscape and genetic variables.

Deriving pairwise connectivity metrics for species without relevant field data (that is, within the study area) such as badgers would have had considerable uncertainty given that substantial individual and population variation in demography and habitat associations across species' ranges. Even in well-studied species, landscape genetic patterns vary between landscapes, primarily due to the presence or spatial arrangement of important factors (for example, Short Bull *et al.*, 2011). RDA and spatially lagged regressions, in contrast, do not necessitate extrapolation to produce pairwise connectivity metrics and can still detect subtle impacts on gene flow as seen with Agriculture in badgers (Kierepka and Latch, 2015). Therefore, their flexibility in variable type makes ordination 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, ordination 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; Fortin and Legendre, 2010; Kierepka and Latch, 2015). 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 (Fortin and Legendre, 2010; Kierepka and Latch, 2015), ordination and regression techniques likely would perform well in situations with complex patterns in gene flow (for example, Bowen *et al.*, 2005; Kamler *et al.*, 2013) and biased sampling. Although ordination 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 (Kierepka and Latch, 2015). When combined with

simulations to quantify potential errors, ordination and regression techniques can be a powerful tool to inform conservation and management efforts in poorly understood species.

#### CONCLUSIONS

Overall, this study provides important information for future management of the American badger in Wisconsin, a protected population that is genetically distinct due to the Mississippi River and Great Lakes (Kierepka, 2014). 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 (that is, conduits of dispersal or density-dependent dispersal). This study demonstrates the utility of ordination and regression methods within individual-based landscape genetics, which to date, were largely restricted to population-based investigations of well-studied species. With ordination 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.

#### DATA ARCHIVING

Data for this study (R code, genotype and landscape data) are available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.mr909>

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank the Wisconsin DNR (Funding from WDNR Federal Wildlife Restoration Grant W-160-P) and the University of Wisconsin-Milwaukee Graduate School (Distinguished Dissertation Fellowship) for financial support. This project would not have been possible without the hundreds of DNR personnel and Wisconsin citizens who reported badger activity in Wisconsin. In particular, D Sample, J Olson and B Sloss were critical to the success of the Wisconsin badger project. We also extend gratitude to the British Columbia badger project for providing hair snare design advice and the Milwaukee County Zoo's staff and resident badger, Tink, for testing hair snare prototypes. Laboratory assistance provided by T Burke, AC Frantz and K McKenney is greatly appreciated. S Robinson and M Johansson provided additional support with R coding of spatially lagged regressions and spatial autocorrelations, respectively. We also thank the anonymous reviewers for providing critical insight that greatly improved this manuscript.

- 
- Anderson CD, Epperson BK, Fortin MJ, Holderegger R, James PMA, Rosenberg MS *et al.* (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol Ecol* **19**: 3565–3575.
- Apps CD, Newhouse NJ, Kinley TA (2002). Habitat associations of American badgers in southeastern British Columbia. *Can J Zool* **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. *Ann Stat* **29**: 1165–1188.
- Bivand R, Altman M, Anselin L, Assunção R, Berke O, Bernat A *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 13 March 2013).
- Blanchong JA, Samuel MD, Scribner KT, Weckworth BV, Langenberg JA, Filcek KB (2008). Landscape genetics and spatial distribution of chronic wasting disease. *Biol Lett* **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*). *Mol Ecol* **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*). *Land Ecol* **21**: 877–889.
- Burnham KP, Anderson DR (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Verlag: New York, NY, USA.
- Carlsson J (2008). Effects of microsatellite null alleles on assignment testing. *J Hered* **99**: 616–623.
- 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*) (Carnivora: Mustelidae). *Mol Ecol 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. *Conserv Genet* **7**: 197–211.
- Chapuis M, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* **24**: 621–631.
- Clark R, Brown WS, Stechert R, Zamudio KR (2010). Roads, interrupted dispersal, and genetic diversity in timber rattlesnakes. *Conserv Biol* **24**: 1059–1069.
- Clark GP, White PCL, Harris S (1998). Effects of roads on badger *Meles meles* populations in south-west England. *Biol Conserv* **86**: 117–124.
- Clevenger AP, Wierzchowski J, Chruszcz B, Gunson K (2002). GIS-generated, expert-based models for identifying wildlife habitat linkages and planning mitigation passages. *Conserv Biol* **16**: 503–514.
- 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. *Can J Zool* **88**: 536–545.
- Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *Am Nat* **168**: 486–499.
- Cushman SA, Landguth EL (2010). Spurious correlations and inference in landscape genetics. *Mol Ecol* **19**: 3592–3602.
- Cushman SA, Lewis JS (2010). Movement behavior explains genetic differentiation in American black bear. *Land Ecol* **25**: 1613–1625.
- Dakin EE, Avise JC (2004). Microsatellite null alleles in parentage analysis. *Heredity (Edinb)* **93**: 504–509.
- Davis CS, Strobeck C (1998). Isolation, variability, and cross-species amplification of polymorphic micro-satellite loci in the family Mustelidae. *Mol Ecol* **7**: 1776–1778.
- Dharmarajan G, Beatty WS, Rhodes OE Jr (2013). Heterozygote deficiencies caused by a Wahlund effect: dispelling unfounded expectations. *J Wildl Manage* **77**: 226–234.
- Dickson BG, Jenness JS, Beier P (2005). Influence of vegetation, topography, and roads on cougar movement in southern California. *J Wildl Manage* **69**: 264–276.
- Dixo M, Metzger JP, Morgante JS, Zamudio KR (2009). Habitat fragmentation reduces genetic diversity and connectivity among top populations in the Brazilian Atlantic Coastal Forest. *Biol Conserv* **142**: 1560–1569.
- Domingo-Roura X, Macdonald DW, Roy MS, Marmi J, Terradas J, Woodroffe R *et al.* (2003). Confirmation of low genetic diversity and multiple breeding females in a social group of Eurasian badgers from microsatellite and field data. *Mol Ecol* **12**: 533–539.
- Duffy AJ, Landa A, O'Connell M, Stratton C, Wright JM (1998). Four polymorphic microsatellites in wolverine, *Gluo gulo*. *Anim Genet* **29**: 63.
- Duquette JF, Gehrt SD (2014). Badger (*Taxidea taxus*) resource selection and spatial ecology in intensive agricultural landscapes. *Am Midl Nat* **171**: 116–127.
- Epperson BK, McRae BH, Scribner K, Cushman SA, Rosenberg MS, Fortin M *et al.* (2010). Utility of computer simulations in landscape genetics. *Mol Ecol* **19**: 3549–3564.
- Epps CW, Palsbøll PJ, Wehausen JD, Roderick GK, Ramey RR II, McCullough DR (2005). Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecol Letters* **8**: 1029–1038.
- Ernest HB, Vickers TW, Morrison SA, Buchalski MR, Boyce WM (2014). Fractured genetic connectivity threatens a southern California puma (*Puma concolor*) population. *PLOS One* **9**: e107985.
- Evanno G, Regnaut S, Goudet J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Fleming MA, Ostrander EA, Cook JA (2002). Microsatellite markers for American mink (*Mustela vison*) and ermine (*Mustela erminea*). *Mol Ecol* **8**: 1352–1355.
- Fortin 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. *Mol Ecol Resour* **10**: 831–844.
- Frantz AC, Bertouille S, Eloy MC, Licoppe A, Chaumont F, Flamand MC (2012). Comparative landscape genetic analyses show a Belgian motorway to be a gene flow barrier for red deer (*Cervus elaphus*), but not wild boars (*Sus scrofa*). *Mol Ecol* **21**: 3445–3457.
- 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? *J Appl Ecol* **46**: 493–505.
- Frantz AC, Pope LC, Carpenter PJ, Roper TJ, Wilson GJ, Delahay RJ *et al.* (2003). Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Mol Ecol* **12**: 1649–1661.
- Frantz AC, Linh San ED, Pope LC, Burke T (2010a). Using genetic methods to investigate dispersal in two badger (*Meles meles*) populations with different ecological characteristics. *Heredity* **104**: 493–501.
- Frantz AC, Pope LC, Etherington TR, Wilson GJ, Burke T (2010b). Using isolation-by-distance-based approaches to assess the barrier effect of linear landscape elements on badger (*Meles meles*) dispersal. *Mol Ecol* **19**: 1663–1674.
- Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L *et al.* (2011). Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogramm Eng Rem S* **77**: 858–864.
- Galpern P, Manseau M, Wilson P (2012). Grains of connectivity: analysis at multiple scales in landscape genetics. *Mol Ecol* **21**: 3996–4009.
- Garraway CJ, Bowman J, Wilson PJ (2011). Using a genetic network to parameterize a landscape resistance surface for fishers, *Martes pennanti*. *Mol Ecol* **20**: 3978–3988.
- Goldberg CS, Waits LP (2010). Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. *Mol Ecol* **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. *Mol Ecol* **22**: 3888–3903.
- Greenwood PJ (1980). Mating systems, philopatry and dispersal in birds and mammals. *Anim Behav* **28**: 1140–1162.
- Guillot G, Rousset F (2013). Dismantling the Mantel tests. *Methods Ecol Evol* **4**: 336–344.
- Hardy OJ, Vekemans X (2002). SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol 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.
- Jombart T, Devillard S, Dufour A-B, 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. *Ecol Evol* **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*). *Can J Zool* **90**: 1231–1242.
- Kierepka EM, Latch EK (2015). Performance of partial statistics in individual-based landscape genetics. *Mol Ecol Resour* **15**: 512–525.
- Kierepka EM (2014). *Landscape genetics of the American badger: understanding challenges with elusive species*. Ph.D. thesis. University of Wisconsin-Milwaukee: Milwaukee, WI, USA.
- Kitchell J (2008). *Wisconsin wildlife harvest summary 1930–2007*. Wisconsin Department of Natural Resources: Madison, Wisconsin, USA.
- Landguth EL, Cushman SA (2009). CDPOP: a spatially explicit cost distance population genetics program. *Mol Ecol Resour* **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. *Mol Ecol* **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.
- Messick JP, Hornocker MG (1981). Ecology of the badger in southwestern Idaho. *Wildlife Monogr* **76**: 1–53.
- Ministry of Environment Ecosystems Branch for the Resources Information Standards Committee (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. Victoria, BC, Canada.
- Navidi W, Arnheim N, Waterman MS (1992). A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *Am J Hum Genet* **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 15 February 2013).
- Omernik JM (1987). Ecoregions of the conterminous United States. Map (scale 1:7,500,000). *Ann Assoc Am Geogr* **77**: 118–125.
- Oyler-McCance SJ, Fedy BC, Landguth EL (2013). Sample design effects in landscape genetics. *Conserv Genet* **14**: 275–285.
- Pearce JK, Cherry K, Drielsma M, Ferrier S, Whish G (2001). Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. *J Appl Ecol* **38**: 412–424.
- Pope LC, Domingo-Roura X, Erven K, Burke T (2006). Isolation by distance and gene flow in the Eurasian badger (*Meles meles*) at both a local and broad scale. *Mol Ecol* **15**: 371–386.
- 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 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 2 January 2013).
- Raymond M, Rousset F (1995). GENEPOP (version 1.2): population-genetics software for exact tests and ecumenicism. *J Hered* **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*). *Mol Ecol* **21**: 3078–3093.
- Rizkalla CE, Swihart RK (2007). Explaining movement decisions of forest rodents in fragmented landscapes. *Biol Conserv* **140**: 339–348.
- 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. *Mol Ecol* **21**: 4190–4205.
- Rosenberg DK, Noon BR, Megahan JW, Meslow EC (1998). Compensatory behavior of *Ensatina eschscholtzii* in biological corridors: a field experiment. *Can J Zool* **76**: 117–133.
- Rousset F (2000). Genetic differentiation between individuals. *J Evol Biol* **13**: 58–62.
- Row JR, Wilson PJ, Gomez C, Koen EL, Bowman J, Thornton D *et al.* (2014). The subtle role of climate change on population genetic structure in Canadian lynx. *Glob Change Biol* **20**: 2076–2086.
- Schultz CB, Crone EE (2001). Edge-mediated dispersal behavior in a prairie butterfly. *Ecology* **82**: 1879–1892.
- Schwartz MK, McKelvey KS (2009). Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conserv Genet* **10**: 441–452.
- Schwartz MK, Copeland JP, Anderson NJ, Squires JR, Inman RM, McKelvey KS *et al.* (2009). Wolverine gene flow across a narrow climatic niche. *Ecology* **90**: 3222–3232.
- Segelbacher G, Cushman SA, Epperson BK, Fortin MJ, Francois O, Hardy OJ *et al.* (2010). Applications of landscape genetics in conservation biology: concepts and challenges. *Conserv Genet* **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 *et al.* (2011). Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Mol Ecol* **20**: 1092–1107.
- Sleeman DP, Davenport J, More SJ, Clegg TA, Griffin JM, O'Boyle I (2009). The effectiveness of barriers to badger *Meles meles* immigration in the Irish Four Area Project. *Eur J Wildl Res* **55**: 267–278.
- Spear SF, Balkenhol N, Fortin M, McRae BH, Scribner K (2010). Use of resistance surfaces for landscape genetics: considerations for parameterization and analysis. *Mol Ecol* **19**: 3576–3591.
- Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson CS, Robinson S *et al.* (2007). Putting the 'landscape' in landscape genetics. *Heredity* **98**: 128–142.
- Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N *et al.* (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res* **24**: 3189–3194.
- van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* **4**: 535–538.
- 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)*. Wisconsin Department of Natural Resources: Madison, Wisconsin, USA.
- Wright S (1943). Isolation by distance. *Genetics* **28**: 114–138.
- Wydeven AP, Wiedenhoef JE, Dhuey B (1998). *Status of Badgers in Wisconsin 1987–1998*. Wisconsin Department of Natural Resources: Madison, Wisconsin, USA.
- Zalewski A, Piertney SB, Zalewska H, Lambin X (2009). Landscape barriers reduce gene flow in an invasive carnivore: geographical and local genetic structure of American mink in Scotland. *Mol Ecol* **18**: 1601–1615.
- Zeller KA, McGarigal K, Whiteley AR (2012). Estimating landscape resistance to movement: a review. *Landscape Ecol* **27**: 777–797.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)