

2015

# Assessing the permeability of landscape features to animal movement: Using genetic structure to infer functional connectivity

Sarah J. Anderson

*Minnesota State University Moorhead*

Elizabeth M. Kierepka

*University of Wisconsin Milwaukee*

Robert K. Swihart

*Purdue University, [rswhart@purdue.edu](mailto:rswhart@purdue.edu)*

Emily K. Latch

*University of Wisconsin Milwaukee*

Follow this and additional works at: <http://docs.lib.purdue.edu/fnrpubs>



Part of the [Forest Sciences Commons](#), and the [Population Biology Commons](#)

---

## Recommended Citation

Anderson S.J., Kierepka E.M., Swihart R.K., Latch E.K., Rhodes O.E., Jr. Assessing the permeability of landscape features to animal movement: Using genetic structure to infer functional connectivity. PLOS ONE Volume 10, Issue 2, 26 February 2015, Article number e0117500. DOI: 10.1371/journal.pone.0117500

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

RESEARCH ARTICLE

# Assessing the Permeability of Landscape Features to Animal Movement: Using Genetic Structure to Infer Functional Connectivity

Sara J. Anderson<sup>1,2</sup>, Elizabeth M. Kierepka<sup>3\*</sup>, Robert K. Swihart<sup>2</sup>, Emily K. Latch<sup>3</sup>, Olin E. Rhodes, Jr.<sup>2,4</sup>

**1** Biosciences Department, Minnesota State University Moorhead, 1104 7<sup>th</sup> Ave, Moorhead, MN, 56563, United States of America, **2** Department of Forestry and Natural Resources, 715 W. State Street, Purdue University, West Lafayette, IN, 47907, United States of America, **3** Behavioral and Molecular Ecology Group, Department of Biological Sciences, University of Wisconsin-Milwaukee, 3209 N. Maryland Ave., Milwaukee, WI, 53024, United States of America, **4** Savannah River Ecology Laboratory, PO Drawer E, Aiken, SC, 29802, United States of America

\* [Liz.Kierepka@gmail.com](mailto:Liz.Kierepka@gmail.com)



OPEN ACCESS

**Citation:** Anderson SJ, Kierepka EM, Swihart RK, Latch EK, Rhodes O (2015) Assessing the Permeability of Landscape Features to Animal Movement: Using Genetic Structure to Infer Functional Connectivity. PLoS ONE 10(2): e0117500. doi:10.1371/journal.pone.0117500

**Academic Editor:** Rob Slotow, University of KwaZulu-Natal, SOUTH AFRICA

**Received:** August 12, 2014

**Accepted:** December 26, 2014

**Published:** February 26, 2015

**Copyright:** © 2015 Anderson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All files are available from the Dryad database (doi: [10.5061/dryad.p5hd0](https://doi.org/10.5061/dryad.p5hd0)).

**Funding:** Financial support for this work was provided by: 1) Cooperative State Research Education and Extension Service, U.S. Department of Agriculture under Agreement no. 2000-04649 (<http://www.csrees.usda.gov/Extension/>), 2) the John S. Wright Fund (<http://www.treefund.org/home>), 3) Department of Forestry and Natural Resources, Purdue University (<https://ag.purdue.edu/fnr/Pages/default.aspx>), and 4) U.S. Department of Education Graduate Assistance in Areas of National Need

## Abstract

Human-altered environments often challenge native species with a complex spatial distribution of resources. Hostile landscape features can inhibit animal movement (i.e., genetic exchange), while other landscape attributes facilitate gene flow. The genetic attributes of organisms inhabiting such complex environments can reveal the legacy of their movements through the landscape. Thus, by evaluating landscape attributes within the context of genetic connectivity of organisms within the landscape, we can elucidate how a species has coped with the enhanced complexity of human altered environments. In this research, we utilized genetic data from eastern chipmunks (*Tamias striatus*) in conjunction with spatially explicit habitat attribute data to evaluate the realized permeability of various landscape elements in a fragmented agricultural ecosystem. To accomplish this we 1) used logistic regression to evaluate whether land cover attributes were most often associated with the matrix between or habitat within genetically identified populations across the landscape, and 2) utilized spatially explicit habitat attribute data to predict genetically-derived Bayesian probabilities of population membership of individual chipmunks in an agricultural ecosystem. Consistency between the results of the two approaches with regard to facilitators and inhibitors of gene flow in the landscape indicate that this is a promising new way to utilize both landscape and genetic data to gain a deeper understanding of human-altered ecosystems.

## Introduction

Fragmentation of natural landscapes by anthropogenic attributes, such as agriculture, urbanization, and transportation infrastructure, alters the manner in which animals utilize remnant native habitats [1], [2]. In fragmented agricultural ecosystems, native habitats are generally reduced in area, resulting in a reconfiguration and loss of connectivity between native habitats.

Award P200A030188 (<http://www2.ed.gov/programs/gaann/index.html>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

The formerly continuous habitats may then be reduced to small patches separated by potentially inhospitable matrix habitat [3], [4]. As a result of the redistribution and loss of connectivity of habitat elements, successful movements of individuals between patches can be markedly reduced [5], either by direct mortality or reluctance to leave the safety of a known environment [6]. Over time, wildlife populations in fragmented landscapes can become genetically differentiated and may lose genetic diversity due to isolation and the acceleration of genetic drift [7–9].

Of particular concern in fragmented ecosystems is the concept that a long-term reduction in genetic diversity and impediments to gene flow can inhibit future adaptation of species [10]. However, if matrix habitats do not completely impede movement between patches, species inhabiting fragmented landscapes may function as a metapopulation. Metapopulations are typified by low levels of dispersal between populations and patterns of population extirpation and recolonization wherein populations inhabiting small, isolated patches are at greater risk of extinction than those inhabiting larger or more connected patches [11]. To help understand how to maintain connectivity between isolated patches, one method is to identify habitats that impede dispersal, a process that generally involves detecting barriers to gene flow.

In fragmented ecosystems, the term ‘barrier’ refers to possible impediments to individual movement that exist in the matrix between patches of ideal habitat. When matrix habitats prevent enough movement to cause genetic differentiation between patches, resultant patterns of genetic variation across a landscape should be correlated with matrix habitats. The desire to understand the relationship between landscape heterogeneity and genetic differentiation has largely led to the formation and prolific growth of the field of landscape genetics [12], [13]. The attraction of landscape genetic approaches for fragmented ecosystems is that assumptions about biological processes are more realistic than traditional population genetic models [14]. Importantly, landscape genetic models do not require *a priori* groupings and genetic patterns can be analyzed at the individual scale. Another benefit of using landscape genetics for fragmentation studies is its ability to disentangle multiple influences on gene flow because genetic variation is often influenced by multiple factors that operate at different spatial [15–18] and temporal scales [19], [20].

Landscape genetic models often depend on accurate definition of the functional biological extent of populations, particularly in management and conservation efforts [21], [22]. Unfortunately, defining biologically meaningful populations is especially problematic within the context of fragmented environments [23]. For example, understanding how factors like the spatial distribution of habitat attributes and the composition of the intervening matrix habitats contribute to the spatial extent of population structure is rarely intuitive. Fortunately, Bayesian approaches provide a method to use individual-based data for detection of underlying population structure at various spatial scales. Bayesian clustering algorithms have proven to be particularly useful for identifying major barriers to gene flow in a variety of species, often confirming expectations that topographic (e.g., [24–26]) and anthropogenic barriers (e.g., [27], [28]) limit the ability of animals to move freely through their environments. While these approaches have been extensively used to elucidate population subdivision resulting from prominent habitat features, Bayesian algorithms can also aid in evaluating how more cryptic factors impact gene flow in heterogeneous landscapes [29–31].

With sufficiently dense, spatially explicit genetic and landscape attribute data, we should be able to use Bayesian clustering tools to not only hypothesize about the presence of cryptic barriers to gene flow, but also to explicitly identify those habitat attributes contributing most significantly to landscape level permeability, even in complex landscapes with relatively subtle changes in fine-scale habitat attributes. One method to evaluate the effects of fragmentation is to utilize individual-based estimates of ancestry or assignment relating to each genetic cluster (i.e., *q*-values or posterior probabilities) inferred from Bayesian algorithms within landscape

genetic models. These genetic variables can act as response variables in statistics that evaluate how landscape parameters influence genetic variation. For example, several studies have utilized genetic surfaces based on ancestry coefficients from Bayesian clustering programs to investigate how specific landscape features (i.e., unsuitable habitats, roads, and topographic barriers) impacted gene flow [18], [32]. Spatially explicit Bayesian algorithms (e.g., Geneland, [33], [34]) may be particularly suited to identifying the effects of fragmentation because they include spatial coordinates as priors, and thus, can infer spatially explicit boundary areas and, by extension, spatial isolating features between biologically relevant populations in the landscape (e.g., [27], [35], [36]).

The main goal of this study, therefore, was to assess the utility of spatially explicit Bayesian methods for detection of fine-scale habitat attributes that inhibit gene flow, leading to evidence of population subdivision, in cryptic, fragmented environments. To accomplish this objective we focused our research in the highly fragmented, human-dominated, agricultural landscape of northern Indiana. Our study species within this landscape was the eastern chipmunk (*Tamias striatus*), a species that exhibits substantial evidence of fine-scale genetic structure within our study area and for which a suite of 12 highly polymorphic microsatellites has been developed [37]. Chipmunks generally select forested habitats or those with adequate tree cover, so movement (and gene flow) was expected to be facilitated by forest and impeded by unsuitable habitats such as agriculture and roads. This study utilized spatially explicit Bayesian clustering to derive population structure from genotypic and spatial coordinate data for 1,422 eastern chipmunks distributed across 33 distinct 23 km<sup>2</sup> study cells in northern Indiana to address two main objectives. First, we examined if the intervening habitat between pairs of patches with and without gene flow were characterized by different fine-scale habitat attributes. Second, we analyzed whether fine-scale land cover features could predict the probability of population membership calculated via spatially explicit Bayesian clustering methods.

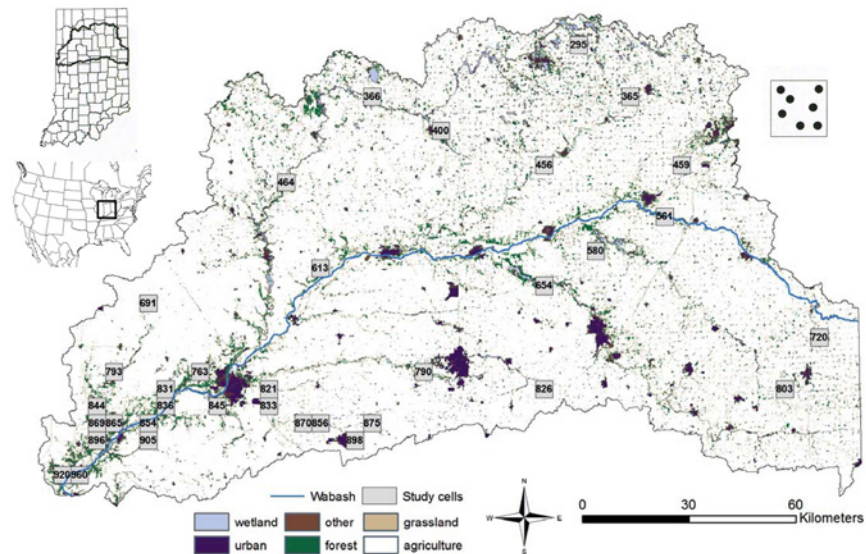
## Methods

### Study Area and Sampling

Our study area was located in north-central Indiana, within the upper Wabash River basin (UWB; Fig. 1). The UWB drains over 20% (>20,000 km<sup>2</sup>) of the state [38] and contains eight major watersheds. Extensive land use change has occurred since the arrival of European settlers in the 17th century, primarily from forest to agriculture. Prior to European settlement, forest cover statewide is estimated to have been 87% [38], [39]. In contrast, forest cover during sample collection in UWB was about 8% (Fig. 1), compared to 19% statewide. Remaining native forests (predominantly oak-hickory-maple [*Quercus-Carya-Acer*]) in the basin are highly fragmented with the largest tracts confined to major drainages where floodplains or locally steep topography prohibits agriculture [38]. Possible corridors are dispersed throughout the region in various forms, such as small woodlots, fence rows, streams, shrubland, and windbreaks. The majority of land area was privately owned (96%) with 88% of the area designated as agriculture.

### Sample Collection

Eastern chipmunks were trapped at multiple sampling sites within 5, 18, and 12 of the 35, 23km<sup>2</sup> study cells from late May to early August in 2001, 2002, 2003, respectively. Prior to each field season, all 30 X 30 m pixels in each cell were classified into 1 of 5 land cover categories (agriculture, forest, grassland, wetland, or urban), and then sampled according to a stratified-random design. To capture the variation within the study area, trapping grids were placed in sampling sites based on land cover so that natural land cover types (i.e., grassland, forest, and



**Fig 1. Map of study area in northern Indiana showing distribution of study cells.** The blue line corresponds to the Wabash River, and land cover is colored. Right inset demonstrates several sample sites within each study cell. Left inset shows the study area location in North America and Indiana with counties outlined. All study cells are labeled by their ID.

doi:10.1371/journal.pone.0117500.g001

wetland; 27.8% of grids each) were disproportionately represented as compared to urban and agriculture (13.9 and 2.8% of all grids respectively). We randomly chose sampling sites (i.e., where trapping grids were placed) within patches of habitat by randomly selecting pixels of an appropriate habitat according to the predefined proportions stated above within each study cell ( $n = 35$ ). Forest site selection had an additional step because they were also stratified according to forest patch size (small  $< 5$  ha; medium 5–50 ha; large  $> 50$  ha), and then were selected according to their  $\log_{10}$  area within the three size categories. In total, a maximum of 45 sampling sites were selected per study cell in a summer [38], [40], and each forest habitat patch generally had 1–3 sampling sites [38]. A full description of study cell and sampling site selection is described in [38].

Each sampling site within a cell contained a grid of Fitch live traps (2001), Sherman live traps (2003), or a mixture of both (2002) spaced 15 m apart. Grid dimensions varied by trap year and patch size. Grids in 2001 were primarily 3 traps x 3 traps, with a few 7x7 grids where forest patch sizes were large enough to accommodate this arrangement. In 2002 and 2003, 5x5 grids were used whenever space allowed; otherwise 4x4 or 3x3 grids were used. If a forest patch was not large enough for a 3x3 grid, it was not sampled. Habitat corridors were identified as treed and non-treed land cover features less than 30 m wide and were fitted with 5x2 grids. Simultaneous trapping for raccoons was conducted at the edges of the trap grids to limit disturbance by raccoons to the small mammal traps.

Traps were baited with black-oil sunflower seeds. A pre-bait period with traps locked open occurred for 3 days, followed by a 5-day trap-check session, during which traps were checked twice daily. Each site had one trapping session. All animals were handled according to procedures approved by the Purdue Animal Care and Use Committee under protocol #01–024. Ear clips were taken from each individual using sterile scissors, and treated with ferric subsulfate in cases of excess bleeding. Animals were then released, and all tissues were stored in at  $-80^{\circ}\text{C}$  prior to DNA extraction.



## DNA Extraction and Genotyping

We extracted DNA from ear tissue using an ammonium acetate protein precipitation protocol (modified form the PUREGENE kit; Gentra Systems) followed by an ethanol wash. Genomic DNA was run on 2% agarose gels stained with ethidium bromide to determine quality and quantity of DNA samples prior to genotyping. Amplification of twelve microsatellite loci [37] by multiplex PCR took place in 10  $\mu$ L reactions volumes with 20 ng of template DNA, 0.2 mM of each dNTP, 1 U of *Taq* DNA polymerase (NEB) and 2X Thermopol reaction buffer (20 mM Tris-HCl, 10 mM  $(\text{NH}_4)_2\text{SO}_4$ , 10 mM KCl, 2 mM  $\text{MgSO}_4$ , 0.1% Triton X-100; NEB). Primer concentrations were adjusted so that intensities of the final products per multiplex reaction were approximately equal [37]. The amplification conditions included an initial denaturation step at 94°C for 2 min, then 35 cycles of 94°C for 30 s, annealing temperature [37] for 30 s, 72°C for 30 s, then a final extension of 72°C for 10 min and a soak at 60°C for 45 min. The PCR products were sized on an Applied Biosystems 3730 automated sequencer, and the genotypes were determined for all loci in all individuals using the software GeneMapper 3.7 (Applied Biosystems).

Several quality control measures were used to confirm the accuracy of genotypes. A negative control, two pre-amplified positive controls, and a concurrently amplified positive control were run on every 96-well plate. To standardize allele calling, ninety-two samples from each multiplex set were amplified twice, scored independently by two individuals, and examined for inconsistencies. Additionally, any ambiguous samples or samples with low quality flags according to GeneMapper 3.7 were re-amplified and genotyped again at all loci. Any individuals missing one or more genotypes were re-amplified in the multiplex reaction up to two times (the last using undiluted genomic DNA) in an attempt to obtain the genotype, which has the added benefit of confirming previous genotypes at the other loci in the multiplex. If there were still missing genotypes after re-amplifying the multiplex, we used single locus reactions to attempt to retrieve the missing genotypes.

Prior to statistical analyses, we used Cervus 2.0 [41] to calculate null allele frequencies as well as to identify and remove duplicate multilocus genotypes from the dataset. We tested for linkage disequilibrium and Hardy-Weinberg disequilibrium in Genepop 3.4 [42], [43]. To obtain standard errors  $<0.01$ , we increased default parameters to the following: 10,000 dememorization iterations, 2,000 batches, and 10,000 iterations per batch.  $F_{IS}$  values and 95% confidence intervals were calculated after 1000 permutations in R [44] using the package *diversity* (function “divBasic”; [45]). Finally, we tested for isolation-by-distance (IBD) across the entire study area using a simple Mantel test in the R package *vegan* [46] via the “mantel” function.

## Statistical Analysis

To investigate the influence of landscape attributes on the creation of genetic structure in eastern chipmunks, we utilized a hierarchical set of analyses involving five key steps. Each step is described in detail in the following sections. Briefly these were: 1) Conduct genetic-based estimation of populations to delineate population boundaries; 2) Define segments within and between populations to represent potential movement pathways of individuals; 3) Measure the proportion of each land cover type and the average probability of population membership in a series of segment widths to determine the appropriate sampling bandwidth for this system; 4) Utilize regression models to evaluate the role of fine-scale habitat attributes in genetic structure and the influence of landscape features on probability of population membership; and 5) Conduct model validation to assess whether the regression models are effective in predicting the outcomes for a test dataset.

We obtained genetic clusters (hereafter referred to as populations) using the Geneland package [33], [34] in R to cluster individuals within each of 33 of the 23  $\text{km}^2$  study cells sampled for eastern chipmunks (two of the 35 study cells had too few samples to be analyzed). Genetic

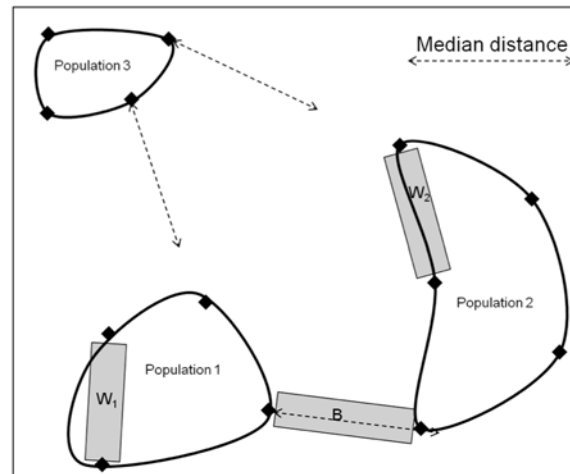
clustering within each study cell was based on multi-locus genotypes with spatial coordinates for all eastern chipmunks captured across all sample sites located within that particular study cell. Bayesian inference is used to estimate the number of populations under the assumption that spatial dependence exists among individuals, so that as geographical distance increases between two individuals, the joint probability that they belong to the same population decreases [34]. The method first estimates the number of populations ( $k$ ), and then calculates assigning the probability of membership of individuals to each of  $k$  populations.

In initial runs of Geneland, we allowed the number of populations within each study cell to vary from 1–6 over 200,000 iterations (thinning = 40; 5,000 iterations retained) of the Markov chain. The number of polygons in the Voronoi tessellation was set close to the number of individuals, as suggested by the authors [34]. Allele frequencies were allowed to be correlated among populations, and an uncertainty radius of 100 m was included on spatial coordinates. The number of populations was estimated by the mode of the posterior distribution of  $k$ . Five independent runs were performed to ensure consistency in the estimation of  $k$ . If there were inconsistencies between runs (e.g., the modal  $k$  or individual assignments were different), we added additional runs with 400,000 iterations (thinning = 80; 5,000 iterations retained). Finally, Geneland may produce empty populations in addition to occupied populations, a phenomenon known as ‘ghost populations’ [34]. The authors suggest ignoring these empty populations and estimating  $k$  only from occupied populations, so we adjusted the estimated  $k$  accordingly.

After  $k$  was estimated, Geneland was run with a fixed  $k$  to obtain individual probabilities of population membership for each individual. The same parameters as when estimating  $k$  were used, except the number of iterations was increased to 400,000 (thinning = 80; 5,000 iterations retained). A grid of 100x100 pixels bound by the outer sample points was created, and the posterior probability of population membership was calculated for each pixel. This process creates a grid of posterior probabilities for each study cell, so study cells with strong genetic structure will have pixels with high posterior probabilities. We used FSTAT 2.9.3 [47] to calculate significance (1,200 randomizations) of pairwise  $F_{ST}$  values among populations of eastern chipmunks identified within each 23 km<sup>2</sup> study cell as well as between study cells to confirm that the populations exhibited significant genetic differentiation.

To classify habitats within each 23 km<sup>2</sup> study cell relative to permeability for gene flow, we first measured pairwise Euclidean distances between all sample sites contained within each population identified by Geneland, excluding those populations comprised of individuals sampled at only one sample point (i.e., no within population distance). Within each study cell, we used distance-based rules to classify line segments connecting sample point as between-population habitat (defined as occurring between two clearly identified populations) versus within-population habitat (defined as occurring within a clearly defined population). If two sample sites belonging to adjacent populations were geographically closer than the median distance of sample sites within those populations, clustering was assumed to be due to a barrier other than simple geographic distance. Pairs of sample sites meeting these criteria were classified as “between” populations, hereafter “B segments”. Alternatively, pairs of sample sites within each of the clearly identified populations on either side of the matrix that most closely matched the between-population distance were classified as “within” population, hereafter “W segments”. To avoid problems arising from spatial autocorrelation associated with the sampling of multiple W segments within any given population, we limited the total number of representative W segment habitats to 1 from each population joined by a B segment (see Fig. 2 for a diagram of this process). Populations that contained only one sample site could not produce W segments; however, this type of population could form one end of a B segment.

Due to the spatial distribution of sample sites within study cells, several (31%) populations were excluded as being associated with clearly defined B segments because the distance



**Fig 2. Diagram of point selection for habitat segments between and within genetically identified populations.** The dashed line indicates the median distance between sample points within a population, averaged across all 70 populations in the study. In this example, the minimum distance between sample points from populations 1 and 2 is less than the median distance and qualified for analysis. However, the minimum distance between sample points from population 3 and the others exceed the median distance, so these did not qualify for a “B” segment. The shaded rectangle represents the segments, B = between populations and W = within populations.

doi:10.1371/journal.pone.0117500.g002

between those populations and neighboring populations was greater than the median distance between sample sites within populations. For these populations, the overall average B segment length (across all sample cells) was calculated (mean = 776 m, median = 760 m) and used to classify 1 W segment within each of these populations if possible. The only populations from which it was not possible to build a W segment were those containing one sample site.

For all segments (B and W), we calculated the proportional area of each of nine habitat types within each segment for a series of segment widths. The widths were varied to create a set of models from which we assessed the sampling bandwidth that best reflected the genetic probabilities of population membership associated with the habitat data. The widths we used were: 3 m, 10 m, 25 m, 50 m, 100 m, 200 m, 400 m, and 1,000 m. The 3 m segment width represents a straight line of minimal width between sample sites. The maximum width of 1,000 m represents a width beyond the maximum movement likely for an eastern chipmunk if an individual traveled along the edge of the defined segment. For each width examined, we recorded the area of each land cover type and the distance between the two sample sites defining the matrix using ArcGIS 9.2 (ESRI). For each segment, we also calculated the average probability of population membership, generated by Geneland, averaging across values for all pixels contained within the segment at each given width.

All defined segments were used to build a series of regression models designed to 1) identify differences in habitat characteristics that might exist between B and W segments of varying widths, and 2) to predict the average probability of population membership (genetic based) from habitat attributes. Eight fine-scale land cover types (forest, non-treed corridors, roads, grassland, shrubs, treed-corridors, urban, and water) served as predictor variables within our regression models. Non-treed corridors were comprised of a mix of grasses, shrubs, and some trees, but mature trees were not the dominant vegetative cover. The water variable included open water bodies as well as wetland areas. We excluded proportion of agriculture from all models due to its high negative correlation with proportion of forest cover. Correlation among the other habitat variables ranged widely across segment widths (S1 Table). Following



calculation of land cover variables, we bootstrap resampled (R function “sample”) our full dataset (B and W segments and their associated land cover variables) 1000 times to calculate confidence intervals around all beta estimates as well as evaluate stochasticity within estimations.

To evaluate which land cover types characterized intervening habitats within or between populations, we used a set of logistic regression models to identify habitat characteristics that differed between B and W segments. The response variable was segment category (between genetically defined populations = 0, within genetically defined populations = 1). Stepwise model selection by AIC was used using the function “stepAIC” in the R package MASS [48]. Stepwise model selection calculates AIC for the full model, and the change in AIC for exclusion of each variable is calculated. If excluding a variable reduces AIC, the next model is evaluated with the remaining subset of variables. Previously excluded variables may be added back to the model if they lower AIC in a later reduced model. The final model is found when addition or exclusion of any variable does not improve the AIC of the model. The final models for each segment width were tested against the null using Chi-Squared tests. Since the same number of B and W segments was used across widths, the null deviance was equal among models as evaluated in [49]. We compared residual deviance among segment widths (i.e., highest residual deviance has the best fit) to determine the best sampling bandwidth for this system.

For our second objective, we used multiple linear regression models with the eight land cover variables as predictor variables to evaluate if land cover attributes can predict the probabilities of population membership calculated from Geneland (both B and W combined). Like the logistic regressions, models were run separately on each bootstrap iteration and segment width (3, 10, 25, 50, 100, 200, 400, and 1000 m). The dependent variable was the average probability of cluster membership across all pixels within each segment at each given width. We used an Anderson-Darling test (function “ad.test”) in the R package nortest [50] to test for normality in our probability data. The Anderson-Darling tests indicated probability values were not normal ( $A = 2.992$ ,  $p < 0.001$ ), so they underwent an arcsine square root transformation to conform to normality. We used the same stepwise selection by AIC as the logistic regression models to determine the best fitting model for each segment width. Significance of the final models for each segment width was assessed (probability of obtaining F-statistic greater than observed for the given degrees of freedom), and adjusted  $R^2$  values were used to assess the segment width that explained the largest proportion of the variance in probability of population membership. All regression calculations were performed in MASS.

To test the predictability of our regression models, we resampled 20% of the individuals within each original bootstrap iteration another 1000 times (total = 1,000,000 iterations per width). For the logistic regression models used to address our first objective, we calculated the probability of getting a B segment from the given composition of habitat and final logistic regression model coefficients for each segment width. A segment was predicted to belong to the W category if its probability was  $> 0.5$ , otherwise the category was B. We then compared the predicted result to the known category for each segment. To assess the strength of the multiple linear regression models for predicting probability of population membership, we calculated the predicted probabilities of population membership and 95% prediction intervals around each probability using the “predict.lm” function in MASS. If an observed probability fell within the 95% confidence intervals, the estimate was considered correct.

## Results

DNA from 1,448 eastern chipmunks was amplified at 12 microsatellite loci, which yielded a total of 17,140 of 17,376 (98.6%) possible genotypes. After duplicate genotypes from possible

recaptures were removed, 1,422 chipmunks remained in the dataset. Test for Hardy-Weinberg and linkage disequilibrium did not indicate any consistent, significant deviations from expected values within or across loci, and null allele frequencies were generally low (0.022–0.123). All diversity metrics (observed and expected heterozygosities, numbers of alleles observed, allelic richness by rarefaction, and  $F_{IS}$  values) were similar between study cells (1-way ANOVA:  $F_{\text{Allelic richness}} = 1.112$ ,  $F_{\text{observed heterozygosity}} = 1.023$ ; all  $p \geq 0.065$ ; [S2 Table](#)). A simple Mantel test revealed significant IBD across the entire study area ( $r = 0.208$ ,  $p = 0.003$ ; [S1 Fig.](#)), and study cells exhibited significant genetic differentiation ( $F_{ST}$  range = 0.012–0.121).

In total, 70 populations were detected in 33 study cells, 8 cells had a single population and 26 had  $k \geq 2$  ([S3 Table](#)). All but one  $F_{ST}$  value between putative populations were significantly different from zero ([S3 Table](#)), supporting that clusters found in Geneland represent true population boundaries. The median distance between sample sites within populations was 1,389 m versus 760 m between populations. In 23 instances, the pairwise distance between sample sites within adjacent populations was less than the median within-population distance. Thus, 23 B segments were created for analyses using these pairs of sample sites. There were 47 instances where pairs of sample sites within each of the clearly identified populations closely matched the average between-population distance. We did not select multiple W segments for two reasons. First, including more W segments would likely introduce high amounts of spatial autocorrelation due to the strong IBD within our study landscapes. Second, we sought to prevent an overrepresentation of W segments within our analyses. In total, 47 W segments (1 W segment per population) and 23 B segments were utilized for our statistical analyses where each segment represented a single data point within our regression models.

Results from logistic regression model selection indicated that the proportions of forest, non-treed corridors, and grassland were consistent discriminatory variables for B or W segment habitats ([Table 1](#)). An increase in any of these variables increased the odds that the segment was within a population. Proportion of treed-corridors remained an important variable after model selection for 100 m, 200 m, and 400 m widths, and an increase in treed-corridors was associated with B segment habitats. At the matrix width of 1,000 m, only grassland was retained as an important variable. At this segment width, increases in the proportions of grasslands corresponded to W segment habitat. All beta estimates were highly consistent across the 1000 bootstrap iterations within each segment width as evidenced by the small standard errors and narrow 95% confidence intervals around each estimate. Percent deviance explained varied from 13.6–28.5% with 200 m (20.8% explained) and 400 m (28.5% explained) being the highest among the eight segment widths ([Table 1](#)).

Similar to the logistic regression analysis, the multiple linear regression retained forest, non-treed corridors, and grassland in the majority of best fit models ([Table 2](#)). The proportion of grassland within segments was retained in the final models identified for all eight segment widths and was positively related to probability of population membership in all models. Proportions of forest (3–100 m) and non-treed habitat (3–400m) also increased the probability of population membership and were retained in the majority of segment widths ([Table 2](#)). While proportion of water was retained in the majority of segment widths (10–200 m), it had a negative effect on probability of population membership. Standard errors around beta estimations were very small (0.001–0.822) across all widths and no 95% confidence intervals included zero, indicating all included variables impacted probability of membership. (The adjusted  $R^2$  value ranged from 0.074 to 0.206 ([Table 2](#)) with the highest values occurred at 200 and 400 m, and rapidly decreased at matrix widths 1000 m.

Collectively, our regressions suggest that the appropriate scale for assessing landscape permeability of small mammals like eastern chipmunks may be less than 1000 m in this landscape because best fit metrics (residual deviance and adjusted  $R^2$ ) peaked at 200–400 m. Therefore,

**Table 1. Final logistic regression models for habitat prediction of segment category in eastern chipmunks after stepwise selection from full models with response variable 0 = between population segment, 1 = within population segment.**

Width	Variable	Estimate	Std. Error	Upper 95% CI	Lower 95% CI	Pr t  > 0	Residual Deviance	p-value
3m							75.112	0.009
	Intercept	-1.206	0.006	-1.194	-1.218	0.136		
	Forest	2.580	0.005	2.589	2.571	0.014	% Exp	
	NTC	20.999	0.054	21.106	20.892	0.106	14.0	
	Grassland	8.501	0.027	8.554	8.447	0.089		
10m							74.186	0.020
	Intercept	-0.187	0.008	-0.171	-0.204	0.790		
	Forest	1.869	0.006	1.882	1.858	0.061	% Exp	
	Grassland	7.765	0.029	7.820	7.709	0.143	13.6	
	Shrubland	-5.878	0.018	-5.843	-5.912	0.177		
	Water	-6.431	0.022	-6.387	-6.475	0.102		
25m							72.298	0.006
	Intercept	-1.023	0.015	-0.994	-1.052	0.264		
	Forest	2.548	0.008	2.564	2.532	0.022	% Exp	
	NTC	22.487	0.065	22.614	22.360	0.110	17.3	
	Grassland	9.586	0.032	9.649	9.522	0.079		
50m							73.307	0.003
	Intercept	-1.491	0.004	-1.482	-1.499	0.053		
	Forest	2.816	0.005	2.825	2.807	0.009	% Exp	
	NTC	27.276	0.060	27.393	27.159	0.085	16.1	
	Grassland	9.970	0.037	10.043	9.898	0.062		
100m							68.377	0.002
	Intercept	-0.753	0.007	-0.739	-0.767	0.376		
	Forest	2.190	0.005	2.201	2.179	0.066	% Exp	
	NTC	26.597	0.099	26.792	26.402	0.136	21.8	
	Grassland	19.122	0.043	19.206	19.038	0.024		
	TC	-24.873	0.063	-24.749	-24.996	0.062		
200m							62.830	<0.001
	Intercept	-0.484	0.008	-0.469	-0.499	0.544		
	Forest	2.265	0.005	2.274	2.256	0.111	% Exp	
	NTC	39.698	0.112	39.918	39.478	0.072	28.1	
	Grassland	21.282	0.037	21.354	21.209	0.014		
	TC	-36.500	0.092	-36.349	-36.710	0.026		
	Shrubland	-7.515	0.021	-7.474	-7.555	0.218		
	Water	-7.015	0.049	-6.918	-7.112	0.119		
400m							62.455	<0.001
	Intercept	-0.520	0.006	-0.507	-0.532	0.541		
	Forest	2.701	0.005	2.711	2.690	0.085	% Exp	
	NTC	54.038	0.139	54.312	54.765	0.023	28.5	
	Grassland	18.304	0.044	18.390	18.218	0.019		
	TC	-53.261	0.105	-53.057	-53.467	0.010		
	Shrubland	-7.813	0.013	-7.788	-7.838	0.157		
	Water	-6.441	0.029	-6.385	-6.497	0.158		
1000m							74.491	0.007
	Intercept	-0.006	0.014	0.021	-0.033	0.467		
	Grassland	13.146	0.043	13.230	13.063	0.020	% Exp	

(Continued)

Table 1. (Continued)

Width	Variable	Estimate	Std. Error	Upper 95% CI	Lower 95% CI	Pr t  > 0	Residual Deviance	p-value
							14.8	

All metrics were calculated from the 1000 bootstrap iterations of each model, and parameters were included if stepwise AIC retained them in at least 900 bootstrap iterations. Shown for each segment width are: the selected variables, mean beta estimates, standard errors with 95% confidence intervals, and average p-values for each beta estimate ( $Pr|t| > 0$ ). Statistics for the overall models include model residual deviance (null deviance = 87.350 for all models), average percent deviance explained by the model (% Exp), and p-value of the Chi-squared test of significance for the model. Land cover abbreviations are: NTC = non-treed corridor and TC = treed corridors.

doi:10.1371/journal.pone.0117500.t001

validation was conducted on 100, 200, and 400 m for both logistic and multiple linear regressions (1,000,000 subsampled populations, 14,000,000 individuals total for each segment width). Assignment of the correct segment category (B or W) for each validated width occurred in 98.01–99.02% of the validation individuals within the logistic regression (14,000,000 individuals total; Fig. 3). In contrast, prediction intervals for 56.35–58.17% of individuals encompassed the observed values (Fig. 3). The drastic differences between the validations in logistic (98.01–99.02%) and multiple regression models (56.35–58.17%) likely stems from the definition of how a predicted value was deemed correct. Multiple regression prediction intervals were smaller than 0.5, the cut-off for the logistic regression, so the stringency in the 95% prediction intervals may explain the lower performance in the multiple regressions.

## Discussion

Although genetic drift as a consequence of isolation by distance likely contributes to variance among populations of eastern chipmunks in our study area, genetic differentiation clearly is not simply a function of spatial proximity in this landscape. We observed breaks in gene flow at distances of only a few hundred meters, well within the capabilities of chipmunks to traverse and less than the median distance between sample sites within populations. The results of our analyses indicate that probabilities of population membership identified across our study landscape likely reflect fine-scale physical and biological barriers to genetic exchange in this environment. Thus, landscape attributes that influence the successful movement, survival, and reproduction of individuals dictate the spatial distribution of populations across our study area, which is characterized by fine-scale habitat features representing pockets of habitat where gene flow readily occurs and gaps where gene flow is restricted. However, two sources of unexplained variance in our analyses that potentially could limit our ability to interpret the permeability of the various landscape attributes we examined are 1) the lack of data on other biological factors, such as predators and competitors, that may be associated with those attributes, and 2) lack of micro-habitat data (e.g., amount of woody debris, ground cover, or average DBH of trees per patch) which may affect the distribution of eastern chipmunks in the study area.

Results from the logistic and linear regression models indicated that forest, non-treed corridors, and grasslands are habitats that are contributing to gene flow and intrapopulation structure. Both the logistic and linear regression identified forest, non-treed corridors, and grasslands as important drivers of genetic structure across the majority of segment widths (3–400 m and 3–200 m in logistic and multiple regression respectively). We also observed little variation in the positive beta estimates in the 1000 bootstrap iterations, which provides robust evidence for the relationship between the land cover variables and whether segments were

**Table 2. Final multiple regression models for habitat composition prediction of probability of population membership for eastern chipmunks collected in northern Indiana after stepwise selection.**

Width	Variable	Estimate	Std. Error	Upper 95% CI	Lower 95% CI	Pr z  > 0	Adjusted R <sup>2</sup>	Model p-value
3m							0.150	0.028
	Intercept	0.627	0.001	0.629	0.626	<0.001		
	Forest	0.211	0.001	0.213	0.209	0.025		
	NTC	1.650	0.004	1.657	1.643	0.058		
	Grassland	0.812	0.001	0.814	0.809	0.117		
10m							0.127	0.013
	Intercept	0.923	0.001	0.924	0.922	<0.001		
	Forest	0.272	0.001	0.273	0.270	0.024		
	NTC	1.622	0.004	1.630	1.615	0.053		
	Grassland	0.819	0.822	0.817	0.002	0.119		
	Water	-0.840	0.002	-0.835	-0.844	0.086		
25m							0.143	0.009
	Intercept	0.929	0.001	0.931	0.927	<0.001		
	Forest	0.266	0.001	0.268	0.265	0.028		
	NTC	1.846	0.005	1.856	1.837	0.081		
	Grassland	0.837	0.001	0.840	0.834	0.048		
	Water	-0.876	0.003	-0.870	-0.883	0.077		
50m							0.158	0.005
	Intercept	0.943	0.001	0.944	0.942	<0.001		
	Forest	0.219	0.001	0.220	0.218	0.063		
	NTC	1.641	0.023	1.687	1.595	0.099		
	Grassland	1.018	0.001	1.021	1.015	0.012		
	Water	-0.944	0.002	-0.940	-0.948	0.053		
100m							0.147	0.007
	Intercept	0.919	0.001	0.920	0.918	<0.001		
	Forest	0.245	0.001	0.246	0.244	0.040		
	NTC	1.784	0.003	1.790	1.777	0.082		
	Grassland	0.930	0.001	0.932	0.928	0.011		
	Water	-0.758	0.002	-0.755	-0.761	0.109		
200m							0.186	0.004
	Intercept	1.070	0.001	1.072	1.068	<0.001		
	NTC	2.387	0.007	2.400	2.373	0.089		
	Grassland	1.375	0.002	1.379	1.371	0.002		
	TC	-2.793	0.007	-2.779	-2.807	0.027		
	Water	-0.789	0.002	-0.785	-0.793	0.097		
400m							0.206	0.001
	Intercept	1.058	0.001	1.059	1.057	<0.001		
	NTC	3.410	0.009	3.429	3.392	0.014		
	Grassland	1.418	0.002	1.421	1.415	0.001		
	TC	-4.391	0.006	-4.380	-4.402	0.002		
1000m							0.074	0.025
	Intercept	1.068	0.002	1.071	1.064	<0.001		

(Continued)



Table 2. (Continued)

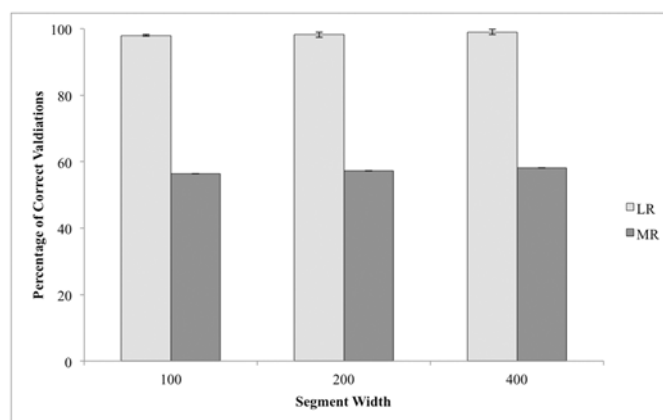
Width	Variable	Estimate	Std. Error	Upper 95% CI	Lower 95% CI	Pr z  > 0	Adjusted R <sup>2</sup>	Model p-value
	Grassland	0.990	0.002	0.993	0.986	0.005		

All metrics were calculated via 1000 bootstrap iterations (average across 1000 iterations or 95% confidence intervals around beta estimates). Each segment width includes the selected variables, coefficient estimates, standard errors, 95% confidence intervals, p-values for each parameter, model adjusted R<sup>2</sup>, and p-value of the overall model. Variables that had an average p-value < 0.05 are given in bold. Note: NTC = non-treed corridors and TC = treed corridors.

doi:10.1371/journal.pone.0117500.t002

B or W. Overall, the consistency of variable selection by both statistical methods and precise beta estimates suggest that forest, non-treed corridors, and grasslands are important to gene flow in this agriculturally dominated landscape.

Forest, as the preferred habitat of eastern chipmunks, was expected to be highly associated with the probability of population membership. Therefore, if a high proportion of the land cover between two points was forest, the area was more likely to be within rather than between populations. Conversely, we expected grassland, a habitat type generally thought to be rather impermeable to small forest dwelling mammals due to its openness and relatively unprotected structure, to be negatively associated with gene flow. For example, [51] reported that eastern chipmunks rarely moved between forest and grassy fields, and [52] claimed that fencerows with only grassy vegetation were inhospitable to chipmunks. However, grassland was the only variable retained in models for all eight segment widths, with more grassland associated with habitat matrix within populations and increased probabilities of population membership. Although not an intuitive result, there is some evidence that eastern chipmunks prefer the forest-grassland ecotone to take advantage of a greater variety of food [53]. Another possible explanation for grassland promoting gene flow is some organisms move through inhospitable habitat quickly to reduce predation risk (e.g., [54–55]). Such compensatory movements [56]



**Fig 3. Proportion of correct validations for 100, 200, and 400m segment widths calculated based on 1,000,000 datasets.** For logistic regressions (light grey), a correct validation occurred when a predicted value of a segment matched the classification of B (0) or W (1) in the observed dataset. A correct validation in the multiple regression (dark grey) was defined as when the observed probability of population membership for a segment fell within the 95% predicted intervals produced from the validation dataset. Logistic regressions were highly accurate with the majority of segments being correctly classified as B or W (98.01–99.02%) while multiple regressions had lower success (56.35–58.17%). Error bars correspond to 95% confidence intervals calculated across 14,000,000 validated segments.

doi:10.1371/journal.pone.0117500.g003

would predict that inhospitable habitats would not act barriers, and could explain why grasslands and non-treed corridors were found to promote gene flow.

Like grassland, non-treed corridors, a land cover variable thought to be unsuitable, also were identified as an important habitat attribute for predicting within population matrix and probability of population membership in both sets of models. A translocation study [57] demonstrated that chipmunks quickly left non-treed corridors after release and moved toward a forest patch, indicating an aversion to such habitat. Although habitat avoidance could suggest that dispersal does not occur in non-treed corridors, the mixture of grasses, shrubs, and trees within such areas may provide an abundance of resources for chipmunks. Thus, eastern chipmunks may find enough protection from predators and the elements in this habitat type to move freely between forest patches, even if individuals have no inclination to remain long-term in such habitat. Vertical stratification (herbs, shrubs, and trees) has been shown to have a positive influence on chipmunk abundance with shrub cover in particular being identified as a critical habitat element for maintenance of chipmunk populations [58]. Indeed, the importance of shrubs as a seasonal food source for chipmunks also has been noted [53], [59], even though dense, short vegetation may inhibit visual range [60].

In terms of the identification of habitat attributes that restrict gene flow of eastern chipmunks in this agricultural landscape, few variables stood out as clear barriers, although two variables of interest were treed corridors and water. Increases in the proportion of treed corridors increased the odds of a habitat being classified as between populations at 100–400m, and was statistically unimportant for predicting the probability of population membership except at 200 and 400 m. This result was very surprising as in this landscape we expected treed corridors to contribute similarly to genetic structure as did forests, since this land cover is simply narrow strips of forest. Instead, treed corridors seemed to function as a weak barrier to gene flow between populations in the handful of models in which it was selected. It is possible that a small number of eastern chipmunks could establish territories in these corridors and reside within them. Chipmunks are highly territorial, so these strips of treed corridors, if occupied, may act as barriers by stopping dispersing individuals and preventing them from reaching other populations. Additionally, predator species may be attracted to treed corridors and use these areas as concentrated feeding grounds, in which case this habitat type could become a population sink. Visual inspection of B segment areas revealed multiple instances of treed corridors occurring perpendicular to the segment between populations, often surrounded by agricultural fields. These perpendicular treed corridors may direct chipmunks to follow horizontally (with respect to the forest patches) for a time, and then turn back to the patch of origin instead of travelling on to a new patch. Alternatively, the association of treed corridors with large tracts of agriculture also may serve to falsely portray the role of treed corridors as isolating mechanisms by masking the underlying effect of agriculture. Further study of the role that treed corridors play in this landscape will be necessary to understand the biological basis for the observed relationship.

Interestingly, we did not detect an effect of roads on gene flow despite roads being implicated as barriers to dispersal for many species [61–64] including eastern chipmunks [62], [65], [66]. For example, [65] found that highways as wide as 90 m were as effective barriers to chipmunk movements as bodies of water twice as wide, and individual chipmunks seem to avoid the mowed roadside verge as well as the road surface regardless of traffic patterns [67]. However, road habitat was not selected as a variable in logistic or multiple regression models in any segment widths. Although we expected to detect a more pronounced effect of roads on gene flow, research also has demonstrated that chipmunks may be less averse to small dirt or gravel roads than to wider paved roads, some of which occur in our study areas [65]. Furthermore, barrier effects of roads combined with intrinsic population characteristics (e.g., high effective

population sizes) may not yield sufficient genetic differentiation to constitute separate populations, even when field studies confirm reduced movement [68], [69]. Thus, while our analyses do not suggest that roads reduce gene flow and increase genetic differentiation between populations, more intensive sampling targeted at forest highway interfaces of differing types may be needed to clarify the specific effects of roads on gene flow in highly fragmented, agricultural landscapes.

One of the primary benefits of landscape genetics for studies of fragmentation is the use of the powerful analytical tools available to population geneticists to evaluate the complex nature of relationships between species ecology and landscape attributes. Bayesian algorithms are extensively used for population delineation in population and landscape genetics, but disentangling fine-scale processes can be difficult to interpret with a broad-scale analysis. In particular, the number of inferred clusters can be overestimated if subtle barriers [70] or isolation-by-distance [71] occur within a study area. Despite potential problems with inferring fine-scale processes via Bayesian clustering algorithms, our approach shows that they can be highly useful in fine-scale analyses even in the presence of confounding variables (e.g., high spatial autocorrelation). For example, outputs from Bayesian algorithms can be utilized to create genetic surfaces (e.g., [17], [32], [72], [73]), which can then further aid in population delineation as well as be used as genetic response variables within landscape genetic statistics (e.g., Mantel tests, regression, or constrained ordination). This study highlights that Bayesian programs can be utilized for many more applications beyond population assignment, and when paired with appropriate statistics like the regressions utilized in this study, can provide powerful evidence of how fragmentation impacts gene flow.

Another advantage of our approach was the ability to utilize regressions to evaluate both the influence of land cover variables on segment category type and the probability of population membership across the study area. Regression models provided the relative contribution of each habitat attribute as a barrier or facilitator of gene flow in the landscape (regression coefficients and measures of variance) and eliminated those that were unimportant for gene flow. Regression coefficients are particularly useful because unlike correlation coefficients in statistics like Mantel tests (see [74]), they can be used to predict how focal landscape variable influence genetic variation (as seen in the validation datasets). Furthermore, our approach suggested that the ideal sampling bandwidth for eastern chipmunks in this landscape was in segments approximately 400 m wide because performance (i.e., model fit statistics) occurred at 400 m wide for both logistic and multiple regression models. Landscape variables can have variable effects according to spatial scales (e.g., [15], [16], [75]), so our method provides a way to disentangle multiple influences on gene flow across spatial scales. Taken together, this approach is highly useful for informing management about how to maintain connectivity in fragmented landscapes because it can predict how landscape variables impact gene flow without *a priori* parameterization and explicitly incorporates spatial scale.

Fine-scale landscape genetic analyses such as the one undertaken in this study have great potential to contribute to conservation and corridor planning because they provide evidence how organisms move in heterogeneous landscapes. In particular, genetic data is often paired with individual-based movement analyses (e.g., [76–78]) that identify paths that organisms are most likely to travel based on costs assigned to different landscape features within a landscape. However, the accuracy of most individual-based modeling approaches rely heavily on the input cost values [79–81], which are generally derived as estimates provided by the investigator [81], with expert opinion and literature-derived information most often serving as the basis for assigning cost values [79], [81]. Several authors [81], [82] have argued that two main faults of individual-based modeling approaches is that we may never know the true cost of dispersal, and that cost values based on expert opinion may be based on the wrong type of data

(e.g., presence/absence of habitat use rather than dispersal cost). If cost values are inaccurate for a given species, the resulting predictions of movement behaviors will be meaningless or misleading, the repercussions of which could lead to ineffective or detrimental conservation and management efforts. As such, the assignment of initial cost values is extremely important, and a pressing question for those who utilize all forms of individual-based modeling is how to most accurately assess cost values for habitats within a landscape.

The concept of using genetic information to inform individual-based analyses began to emerge due to concerns over improperly parameterized cost values. For example, recent studies have employed intense computational methods involving hundreds to thousands of least cost paths with different expert-derived cost values to reveal the parameter set that best matches a set of observed genetic parameters [82–84]. Also, methodologies such as [85] have recently been developed to help optimize and develop unbiased cost values for landscape genetic studies. Our approach represents another potential method to parameterize landscapes without including human bias because no *a priori* assignment of cost values is needed. Instead, regression coefficients derived from models developed in independent landscapes like our study could also be used to parameterize resistance surfaces. Once developed, genetically informed individual-based models could be used for a variety of purposes such as comparison to other parameterization (e.g., expert opinion, presence-absence) or optimization ([84], [85]) methods, mapping possible movement corridors, estimating how functional connectivity may change under future land use alterations, or predicting changes in ecological and genetic parameters associated with changing land use practices.

Elucidating the complex relationships between landscape features and gene flow continues to be important in our human-dominated ecosystems as conservation experts try to plan for future challenges by mitigating isolating factors and improving connectivity of natural populations. In this study, utilizing both spatially explicit genetic and landscape attribute data, we have been successful in identifying landscape elements that affect gene flow within and among eastern chipmunk populations inhabiting a highly fragmented agricultural ecosystem. The results of our analyses are intriguing, given that the estimated permeability of both grasslands and treed corridors to chipmunk movement were counterintuitive and thus, point to the need for further study of how this species actually utilizes these landscape attributes. Ultimately, the approaches outlined herein can be extended to a variety of species and the results utilized to more accurately parameterize individual-based models for application to a wide variety of conservation and management issues.

## Supporting Information

**S1 Table. Correlation coefficients and standard deviation of landscape features found in the segments between populations (below diagonal) and within populations (above diagonal) across the eight matrix widths.** Abbreviations are as follows: ag = agriculture, ntc = non-treed corridors and tc = treed corridors.

(DOCX)

**S2 Table. Microsatellite locus information for eastern chipmunks sampled in northern Indiana during 2001–2003 by study cell.** Sample size (N), number of alleles (Na), observed (Ho) and expected heterozygosity (He), allelic richness ( $A_R$ ), and  $F_{IS}$  with the upper and lower 95% confidence intervals are listed per locus and sample cell. Bold  $F_{IS}$  values indicate a significant deficiency of heterozygotes.

(DOCX)

**S3 Table. Number of genetic clusters estimated from Geneland ( $k$ ) and  $F_{ST}$  for each study cell.** Global  $F_{ST}$  values were calculated for each study cell as well as pair-wise  $F_{ST}$  values

between all putative clusters within a study cell. All but one  $F_{ST}$  between putative clusters were significantly different from zero.

(DOCX)

**S1 Fig. Plot of Euclidean distance versus  $F_{ST}/1 - F_{ST}$  between 33 study cells included in this study.** We recorded a positive relationship between geographic and genetic distance, consistent with IBD ( $r = 0.208$ ,  $p = 0.003$ ).

(TIF)

## Acknowledgments

We would like to thank the private landowners in the upper Wabash River basin, Indiana for allowing sampling to be conducted on their properties. We also extend gratitude to the field crews including field technicians and crew chiefs that collected animals and digitized geographic information layers. J.E. Moore, J. Crick, T. Preuss, L. Connolly, and N. Engbrecht were especially integral in design and coordination of field protocols as well as data collection and management. Finally, we are grateful to the many sources that aided in the GIS methods in this manuscript including the Center for Advanced Applications in GIS at Purdue University, National Land Cover Data, Indiana Unified Watershed Assessment, and the National Water Information System.

## Author Contributions

Conceived and designed the experiments: SJA OER RKS. Performed the experiments: SJA RKS. Analyzed the data: SJA OER EMK. Contributed reagents/materials/analysis tools: SJA RKS OER. Wrote the paper: SJA EMK EKL OER.

## References

1. Andreassen HP, Hertzberg K, Ims RA (1998) Space-use responses to habitat fragmentation and connectivity in the root vole *Microtus oeconomus*. *Ecology* 79: 1223–1235.
2. Gibbs JP (1998) Distribution of woodland amphibians along a forest fragmentation gradient. *Land Ecol* 13: 263–268.
3. Wilcox B, Murphy D (1985) Conservation strategy: the effects of fragmentation on extinction. *Am Nat* 125: 879–887.
4. Andr n H (1994) Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat—a review. *Oikos* 71: 355–366.
5. Schumaker NH (1996) Using landscape indices to predict habitat connectivity. *Ecology* 77: 1210–1225.
6. Heino M, Hanski I (2001) Evolution of migration rate in a spatially realistic metapopulation model. *Am Nat* 157: 495–511. doi: [10.1086/319927](https://doi.org/10.1086/319927) PMID: [18707258](https://pubmed.ncbi.nlm.nih.gov/18707258/)
7. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16: 97–159. PMID: [17246615](https://pubmed.ncbi.nlm.nih.gov/17246615/)
8. Hill JK, Hughes CL, Dytham C, Searle JB (2006) Genetic diversity in butterflies: interactive effects of habitat fragmentation and climate-driven range expansion. *Biol Lett* 2: 152–154. PMID: [17148351](https://pubmed.ncbi.nlm.nih.gov/17148351/)
9. Andre T, Lemes MR, Grogan J, Gribel R (2008) Post—logging loss of genetic diversity in a mahogany (*Swietenia macrophylla* King, Meliaceae) population in Brazilian Amazonia. *For Ecol and Manag* 255: 340–345.
10. Opdam P, Wascher D (2004) Climate change meets habitat fragmentation: linking landscape and biogeographical scale levels in research and conservation. *Biol Conserv* 117: 285–297.
11. Hanski I, Gilpin M (1997) *Metapopulation Biology: Ecology, Genetics, and Evolution* Academic Press. 521 p.
12. Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* 18: 189–197.



13. Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Mol Ecol* 19: 3496–3514. doi: [10.1111/j.1365-294X.2010.04691.x](https://doi.org/10.1111/j.1365-294X.2010.04691.x) PMID: [20723061](https://pubmed.ncbi.nlm.nih.gov/20723061/)
14. Pearse DE, Crandall KA (2004) Beyond F-ST: Analysis of population genetic data for conservation. *Conserv Genet* 5: 585–602.
15. Galpern P, Manseau M, Wilson P (2012) Grains of connectivity analysis at multiple scales in landscape genetics. *Mol Ecol* 21: 3996–4009. doi: [10.1111/j.1365-294X.2012.05677.x](https://doi.org/10.1111/j.1365-294X.2012.05677.x) PMID: [22724394](https://pubmed.ncbi.nlm.nih.gov/22724394/)
16. Dudaniec RY, Rhodes JR, Wilmer JW, Lyons M, Lee KE, et al. (2013) Using multilevel models to identify drivers of landscape-genetic structure among management areas. *Mol Ecol* 22: 3752–3765. doi: [10.1111/mec.12359](https://doi.org/10.1111/mec.12359) PMID: [23730800](https://pubmed.ncbi.nlm.nih.gov/23730800/)
17. Balkenhol N, Holbrook JD, Onorato D, Zager P, White C, et al. (2014) A multi-method approach for analyzing hierarchical genetic structures: a case study with cougars *Puma concolor*. *Ecography* 37: 552–563.
18. Fant JB, Havens K, Keller JM, Radosavljevic A, Yates ED (2014) The influence of contemporary and historic landscape features on the genetic structure of the sand dune endemic, *Cirsium pitcher* (Asteraceae). *Heredity* 112: 519–530. doi: [10.1038/hdy.2013.134](https://doi.org/10.1038/hdy.2013.134) PMID: [24398882](https://pubmed.ncbi.nlm.nih.gov/24398882/)
19. Epps CW, Wasser SK, Keim JL, Mutayoba BM, Brashares JS (2013) Quantifying past and present connectivity illuminates a rapidly changing landscape for the African elephant. *Mol Ecol* 22: 1574–1588. doi: [10.1111/mec.12198](https://doi.org/10.1111/mec.12198) PMID: [23398457](https://pubmed.ncbi.nlm.nih.gov/23398457/)
20. Sharma S, Dutta T, Maldonado JE, Wood TC, Panswar HS, et al. (2013) Forest corridors maintain historical gene flow in a tiger metapopulation in the highlands of central India. *Proc R Soc B*. 280: 20131506. doi: [10.1098/rspb.2013.1506](https://doi.org/10.1098/rspb.2013.1506) PMID: [23902910](https://pubmed.ncbi.nlm.nih.gov/23902910/)
21. Moritz C (1994) Defining evolutionarily-significant-units for conservation. *Trends Ecol Evol* 9: 373–375. doi: [10.1016/0169-5347\(94\)90057-4](https://doi.org/10.1016/0169-5347(94)90057-4) PMID: [21236896](https://pubmed.ncbi.nlm.nih.gov/21236896/)
22. Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends Ecol Evol* 22: 25–33. PMID: [16962204](https://pubmed.ncbi.nlm.nih.gov/16962204/)
23. Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol Ecol* 15: 1419–1439. PMID: [16629801](https://pubmed.ncbi.nlm.nih.gov/16629801/)
24. ošić N, ičanová S, Bryja J, Penezić A, irović D (2013) Do rivers and human-induced habitat fragmentation affect genetic diversity and population structure of the European ground squirrel at the edge of its Pannonian range? *Conserv Genet* 14: 345–354. doi: [10.1186/1471-2164-14-345](https://doi.org/10.1186/1471-2164-14-345) PMID: [23705979](https://pubmed.ncbi.nlm.nih.gov/23705979/)
25. Marin JC, González BA, Poulin E, Casey CS, Johnson WE (2013) The influence of the arid Andean high plateau on the phylogeography and population genetics of guanaco (*Lama guanicoe*) in South America. *Mol Ecol* 22: 463–482. doi: [10.1111/mec.12111](https://doi.org/10.1111/mec.12111) PMID: [23206254](https://pubmed.ncbi.nlm.nih.gov/23206254/)
26. Smitsen PJ, Melville J, Sumner J, Jessop TS (2013). Mountain barriers and river conduits: phylogeographical structure in a large, mobile lizard (Varanidae: *Varanus varius*) from eastern Australia. *J Biogeog* 40: 1729–1740. PMID: [24393903](https://pubmed.ncbi.nlm.nih.gov/24393903/)
27. Frantz AC, Bertouille S, Eloy MC, Licoppe A, Chaumont F, et al. (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. doi: [10.1111/j.1365-294X.2012.05623.x](https://doi.org/10.1111/j.1365-294X.2012.05623.x) PMID: [22607276](https://pubmed.ncbi.nlm.nih.gov/22607276/)
28. Hepenstrick D, Thiel D, Holderegger R, Gugerli F (2012) Genetic discontinuities in roe deer (*Capreolus capreolus*) coincide with fenced transportation infrastructure. *Bas Appl Ecol* 13: 631–638.
29. Latch EK, Boarman WI, Walde A, Fleischer RC (2011) Fine-scale analysis reveals cryptic landscape genetic structure in desert tortoises. *PloS One*. Available at: doi: [10.1371/journal.pone.0029581](https://doi.org/10.1371/journal.pone.0029581) PMID: [22242147](https://pubmed.ncbi.nlm.nih.gov/22242147/)
30. Reding DM, Bronikowski AM, Johnson WE (2012) Pleistocene and ecological effects on continental-scale genetic differentiation in the bobcat (*Lynx rufus*). *Mol Ecol* 21: 3078–3093. doi: [10.1111/j.1365-294X.2012.05595.x](https://doi.org/10.1111/j.1365-294X.2012.05595.x) PMID: [22548482](https://pubmed.ncbi.nlm.nih.gov/22548482/)
31. Stronen AV, Navid EL, Quinn MS, Paquet PC, Bryan HM, et al. (2014). Population genetic structure of gray wolves (*Canis lupus*) in a marine archipelago suggests island-mainland differentiation consistent with dietary niche. *BMC Ecol* 14: 11. Available at: doi: [10.1186/1472-6785-14-11](https://doi.org/10.1186/1472-6785-14-11) PMID: [24915756](https://pubmed.ncbi.nlm.nih.gov/24915756/)
32. Murphy M, Evans J, Cushman S, Storfer A (2008) Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography* 31: 685–697.
33. Guillot G, Estoup A, Mortier F, Cosso JF (2005a) A spatial statistical model for landscape genetics. *Genetics* 170: 1261–1280. PMID: [15520263](https://pubmed.ncbi.nlm.nih.gov/15520263/)
34. Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Mol Ecol Notes* 5: 712–715.

35. Blair C, Weigel DE, Balazik A, Keeley TH, Walker FM, et al. (2012) A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Mol Ecol Resources* 12: 822–833. doi: [10.1111/j.1755-0998.2012.03151.x](https://doi.org/10.1111/j.1755-0998.2012.03151.x) PMID: [22551194](https://pubmed.ncbi.nlm.nih.gov/22551194/)
36. Habel JC, Engler JO, Rödder D, Schmitt T (2012) Landscape genetics of a recent extirpation in a burnet moth species. *Conserv Genet* 13: 247–255.
37. Anderson SJ, Fike JA, Dharmarajan G, Rhodes OE (2007) Characterization of 12 polymorphic microsatellite loci for eastern chipmunks (*Tamias striatus*). *Mol Ecol Notes* 7: 513–515.
38. Moore JE, Swihart RK (2005) Modeling patch occupancy by forest rodents: Incorporating detectability and spatial autocorrelation with hierarchically structured data. *J Wildl Manage* 69: 933–949.
39. Smith WB, Faulkner JL, Powell DS (1994) Forest statistics of the United States, 1992. Metric units. U.S. Forest Service General Technical Report NC-168, St. Paul, Minnesota.
40. Urban NA, Swihart RK (2009) Multiscale perspectives on occupancy of meadow jumping mice in landscapes dominated by agriculture. *J Mammal* 90: 1431–1439.
41. Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* 7: 639–655. PMID: [9633105](https://pubmed.ncbi.nlm.nih.gov/9633105/)
42. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361–372. PMID: [1637966](https://pubmed.ncbi.nlm.nih.gov/1637966/)
43. Raymond M, Rousset F (1995) GENEPOP (VERSION-1.2)—Population-genetics software for exact tests and ecumenicism. *J Hered* 86: 248–249.
44. R Core Team 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0. Available: <http://www.R-project.org/>. Accessed 2014 September.
45. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: An R package for the estimation of population genetics parameters and their associated errors, *Methods Ecol and Evol* 4: 782–788.
46. Oksanen J, Kindt R, Legendre P, O'Hara RB (2008) vegan: community ecology package. Version 1.7–81. Available: <http://cran.r-project.org/>. Accessed 2014 February.
47. Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available: <http://www.unil.ch/izea/software/fstat.html>.
48. Ripley B, Hornik K, Gebhardt A, Firth D (2012) Package “MASS”. Available: <http://cran.r-project.org/web/packages/MASS/MASS.pdf>. Accessed 2014 September.
49. Kutner M, Nachtsheim C, Neter J, Li W (2005) Applied Linear Statistical Models, 5th edn. McGraw-Hill/Irwin, New York.
50. Gross J (2006) nortest: Tests for Normality. R package version 1.0. Available: <http://cran.r-project.org/web/packages/nortest/>. Accessed 2014 September.
51. Anderson SJ (2010) Genetic structure of white-footed mice (*Peromyscus leucopus*) and eastern chipmunks (*Tamias striatus*) in an agriculturally fragmented ecosystem: A landscape genetics approach. Dissertation. Purdue University.
52. Wegner J, Merriam G (1979) Movements by birds and small Mammals between a wood and adjoining farmland habitats. *J Appl Ecol* 16: 349–357.
53. Bennett AF, Henein K, Merriam G (1994) Corridor use and the elements of corridor quality—chipmunks and fencerows in a farmland mosaic. *Biol Conserv* 68: 155–165.
54. Bond BT, Leopold BD, Burger LW Jr., Godwin KD (2001) Movements and home range dynamics of cottontail rabbits in Mississippi. *J Wildl Manage* 65: 1004–1013.
55. Rizkalla CE, Swihart RK (2007) Explaining movement decisions of forest rodents in fragmented landscapes. *Biol Conserv* 140: 339–348.
56. Rosenberg DK, Noon BR, Megahan JW, Meslow EC (1998) Compensatory behavior of *Ensatina eschschooltzii* in biological corridors: a field experiment. *Can J Zool* 76: 117–133.
57. Forsyth DJ, Smith DA (1973) Temporal variability in home ranges of eastern chipmunks (*Tamias striatus*) in a southeastern Ontario woodlot. *Am Midl Nat* 90: 107–117.
58. Rizkalla CE, Swihart RK (2007) Explaining movement decisions of forest rodents in fragmented landscapes. *Biol Conserv* 140: 339–348.
59. Geier AR, Best LB (1980) Habitat selection by small mammals of riparian communities—evaluating effects of habitat alterations. *J Wildl Manage* 44: 16–24.
60. Mahan CG, Yahner RH (1996) Effects of forest fragmentation on burrow-site selection by the eastern chipmunk (*Tamias striatus*). *Am Midl Nat* 136: 352–357.
61. Wilkins KT (1982) Highways as barriers to rodent dispersal. *Southwest Nat* 27: 459–460.

62. Mills LS, Conrey RY (2003) Highways as potential barriers to movement and genetic exchange in small mammals. Final Report, Montana Department of Transportation Research Section.
63. Alexander S, Waters N, Paquet P (2005) Traffic volume and highway permeability for a mammalian community in the Canadian Rocky Mountains. *Can Geograph* 49: 321–331.
64. Bissonette JA, Adair W (2008) Restoring habitat permeability to roaded landscapes with isometrically-scaled wildlife crossings. *Biolog Conserv* 141: 482–488.
65. Oxley DJ, Fenton MB, Carmody GR (1974) Effects of roads on populations of small mammals. *J Appl Ecol* 11: 51–59.
66. McGregor RL, Bender DJ, Fahrig L (2008) Do small mammals avoid roads because of the traffic? *J Appl Ecol* 45: 117–123.
67. Ford AT, Fahrig L (2008) Movement patterns of eastern chipmunks (*Tamias striatus*) near roads. *J Mammal* 89: 895–903.
68. Gauffre B, Estoup A, Bretagnolle V, Cosson JF (2008) Spatial genetic structure of a small rodent in a heterogeneous landscape. *Mol Ecol* 17: 4619–4629. doi: [10.1111/j.1365-294X.2008.03950.x](https://doi.org/10.1111/j.1365-294X.2008.03950.x) PMID: [19140985](https://pubmed.ncbi.nlm.nih.gov/19140985/)
69. Rico A, Kindlmann P, Sedláček F (2009) Can the barrier effect of highways cause genetic subdivision in small mammals? *Acta Theriolog* 54: 297–310.
70. Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conserv Genet* 7: 295–302.
71. 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.
72. Row JR, Blouin-Demers G, Loughheed SC (2010) Habitat distribution influences dispersal and fine-scale genetic population structure of eastern foxsnakes (*Mintonius gloydi*) across a fragmented landscape. *Mol Ecol* 19: 5157–5171. doi: [10.1111/j.1365-294X.2010.04872.x](https://doi.org/10.1111/j.1365-294X.2010.04872.x) PMID: [20977510](https://pubmed.ncbi.nlm.nih.gov/20977510/)
73. Latch EK, Reding DM, Heffelfinger JR, Alcalá-Galván CH, Rhodes OE (2014) Range-wide analysis of genetic structure in a widespread, highly mobile species (*Odocoileus hemionus*) reveals the importance of historical biogeography. *Mol Ecol* 23: 3171–3190. doi: [10.1111/mec.12803](https://doi.org/10.1111/mec.12803) PMID: [24863151](https://pubmed.ncbi.nlm.nih.gov/24863151/)
74. Legendre P, Fortin M (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol Ecol Res* 10: 831–844. doi: [10.1111/j.1755-0998.2010.02866.x](https://doi.org/10.1111/j.1755-0998.2010.02866.x) PMID: [21565094](https://pubmed.ncbi.nlm.nih.gov/21565094/)
75. Angelone S, Kienast F, Holderegger R (2011) Where movement happens: scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography* 34: 714–722.
76. Dijkstra E (1959) A note on problems in connexion with graphs. *Numerische Mathematik* 1: 269–271.
77. McRae BH, Dickson BG, Keitt TM, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* 89: 2712–2724. PMID: [18959309](https://pubmed.ncbi.nlm.nih.gov/18959309/)
78. McRae BH, Hall SA, Beier P, Theobald DM (2012) Where to restore ecological connectivity? Detecting barriers and quantifying restoration benefits. *PloS One* 7: e52604. doi: [10.1371/journal.pone.0052604](https://doi.org/10.1371/journal.pone.0052604) PMID: [23300719](https://pubmed.ncbi.nlm.nih.gov/23300719/)
79. Rayfield B, Fortin M, Fall A (2010) The sensitivity of least-cost habitat graphs to relative cost surface values. *Land Ecol* 25: 519–532.
80. Sawyer SC, Epps CW, Brashares JS (2011) Placing linkages among fragmented habitats: do least-cost models reflect how animals use landscapes? *J Appl Ecol* 48: 668–678.
81. Zeller KA, McGarigal K, Whiteley AR (2012) Estimating landscape resistance to movement: a review. *Land Ecol* 27: 777–797.
82. Epps CW, Wehausen JD, Bleich VC, Torres SG, Brashares JS (2007) Optimizing dispersal and corridor models using landscape genetics. *J Appl Ecol* 44: 714–724.
83. Perez-Espona S, Perez-Barberia FJ, McLeod JE, et al. (2008) Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Mol Ecol* 17: 981–996. doi: [10.1111/j.1365-294X.2007.03629.x](https://doi.org/10.1111/j.1365-294X.2007.03629.x) PMID: [18261043](https://pubmed.ncbi.nlm.nih.gov/18261043/)
84. Wang IJ, Savage WK, Shaffer HB (2009) Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). *Mol Ecol* 18: 1365–1374. doi: [10.1111/j.1365-294X.2009.04122.x](https://doi.org/10.1111/j.1365-294X.2009.04122.x) PMID: [19368644](https://pubmed.ncbi.nlm.nih.gov/19368644/)
85. Peterman WE (2014) ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. bioRxiv doi: [10.1101/007575](https://doi.org/10.1101/007575)