



## Research

**Cite this article:** Draheim HM, Moore JA, Etter D, Winterstein SR, Scribner KT. 2016 Detecting black bear source–sink dynamics using individual-based genetic graphs.

*Proc. R. Soc. B* **283**: 20161002.

<http://dx.doi.org/10.1098/rspb.2016.1002>

Received: 12 May 2016

Accepted: 21 June 2016

**Subject Areas:**

ecology, genetics

**Keywords:**

source–sink dynamics, graph theory, genetic relatedness, black bear, connectivity

**Author for correspondence:**

Hope M. Draheim

e-mail: [hdraheim@gmail.com](mailto:hdraheim@gmail.com)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2016.1002> or via <http://rspb.royalsocietypublishing.org>.

# Detecting black bear source–sink dynamics using individual-based genetic graphs

Hope M. Draheim<sup>1</sup>, Jennifer A. Moore<sup>2</sup>, Dwayne Etter<sup>3</sup>, Scott R. Winterstein<sup>4</sup> and Kim T. Scribner<sup>4</sup>

<sup>1</sup>National Forensic Laboratory, US Fish and Wildlife Service, 1490 E Main Street, Ashland, OR 97520, USA

<sup>2</sup>Biology Department, Grand Valley State University, Allendale, MI 49401, USA

<sup>3</sup>Michigan Department of Natural Resources, Wildlife Division, 8562 E. Stoll Road, East Lansing, MI 48823, USA

<sup>4</sup>Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA

Source–sink dynamics affects population connectivity, spatial genetic structure and population viability for many species. We introduce a novel approach that uses individual-based genetic graphs to identify source–sink areas within a continuously distributed population of black bears (*Ursus americanus*) in the northern lower peninsula (NLP) of Michigan, USA. Black bear harvest samples ( $n = 569$ , from 2002, 2006 and 2010) were genotyped at 12 microsatellite loci and locations were compared across years to identify areas of consistent occupancy over time. We compared graph metrics estimated for a genetic model with metrics from 10 ecological models to identify ecological factors that were associated with sources and sinks. We identified 62 source nodes, 16 of which represent important source areas (net flux  $> 0.7$ ) and 79 sink nodes. Source strength was significantly correlated with bear local harvest density (a proxy for bear density) and habitat suitability. Additionally, resampling simulations showed our approach is robust to potential sampling bias from uneven sample dispersion. Findings demonstrate black bears in the NLP exhibit asymmetric gene flow, and individual-based genetic graphs can characterize source–sink dynamics in continuously distributed species in the absence of discrete habitat patches. Our findings warrant consideration of undetected source–sink dynamics and their implications on harvest management of game species.

## 1. Introduction

When habitat quality varies across heterogeneous landscapes, occupancy and relative abundance can concomitantly vary [1]. As population densities increase in higher-quality habitats, intraspecific competition for resources can force individuals to disperse to lower-density and lower-quality habitats [2]. This phenomenon is commonly referred to as source–sink dynamics [3]. Source areas are characterized by higher-quality habitat and a surplus of individuals that emigrate to other areas. By contrast, sink areas are characterized by lower-quality habitat, have negative population growth rates and require an influx of dispersing individuals from source areas to maintain population persistence [4,5]. Source–sink dynamics fundamentally affect population connectivity and spatial genetic structure of species, and have evolutionary and ecological implications that could influence population viability, species persistence and evolutionary potential [3,6,7]. Furthermore, source–sink dynamics may influence the evolutionary trajectory of species by disrupting local adaptation to low-quality (sink) habitats as directional selection is mitigated by an influx of individuals from high-quality (source) habitats [8–10]. Genetic variation may be lost as alleles present in the source populations, which have higher mean fitness across all populations, tend to be fixed [11]. Thus, detecting sources and sinks is critical for understanding microevolutionary processes that affect populations, and for managing wildlife and conserving at-risk species.

Traditionally, source–sink dynamics have been detected using direct methods including field observations or capture–mark–recapture studies [12]. An alternative method is to apply graph theory to estimate asymmetrical connectivity [13]. Graphs are simple visual representations of real systems [14] that are composed of nodes (representing biological units like individuals, populations or habitat patches) that are characterized by attributes like size or quality. Nodes are connected by edges that represent biological phenomena like dispersal. Edges can be symmetric, with uniform weighting between nodes. Alternatively, asymmetric edges are directionally weighted where, for instance, dispersal from patch  $i$  to patch  $j$  is greater than from patch  $j$  to patch  $i$ , a pattern commonly found in source–sink dynamics [15].

Combining population genetics and graph theory offers a compelling framework [7,12,16] for understanding source–sink dynamics and population connectivity. However, for species that are continuously rather than patchily distributed, use of genetic data to estimate gene flow has been hampered by an inability to delineate populations. Use of individual-based measures of pairwise genetic relatedness allows for investigations of fine-scale genetic structure and dispersal patterns of continuously distributed species without *a priori* delineation of populations [17,18]. Pairwise relatedness represents the degree of shared ancestry, or the probability that two alleles at a locus, one taken at random from each of two individuals, are identical by descent [19].

When applied in a graph context, nodes can represent individuals or populations connected by edges representing shared ancestry [13,20]. Continuously distributed species commonly exhibit a pattern of isolation by distance (IBD), where individuals closer in geographical proximity are more genetically similar [21]. Alternatively, if habitats are highly heterogeneous, habitat quality can more strongly impact genetic structure than distance alone [22]. Differential occupancy in heterogeneous landscapes will ultimately affect patterns of relatedness. For example, within source areas with high recruitment and little immigration, average pairwise relatedness will be higher than in sinks due to accumulation of multi-generational family groups. By contrast, average pairwise relatedness will be lower in sink areas, which are characterized by high rates of immigration of individuals with mixed ancestry [12].

Graph theory also provides metrics that quantify the contributions of individual nodes and edges to overall graph connectivity. When applied to management questions, such metrics can provide a novel and powerful means for managers and conservationists to predict the impact of future policy decisions or land use changes. For example, wildlife managers can predict how increasing regional harvest quotas will impact global population connectivity. A need for landscape genetic methods that move beyond assessing the influences of landscape features on gene flow to predicting the impacts of management decisions and landscape change has recently been recognized [23,24].

To characterize source–sink dynamics, which we define here as asymmetric gene flow, we developed an individual-based graph approach to investigate source–sink dynamics in a population of American black bears (*Ursus americanus*) in the northern lower peninsula (NLP) of Michigan, USA. Black bears have home ranges that can overlap depending on resource availability [25–27] and habitat quality (electronic supplementary material, figure S1). Bears in the NLP are geographically isolated as the population is bounded to the

south by a matrix of unsuitable urban and agricultural habitat, and on all other sides by the Great Lakes; thus the population experiences little to no emigration and immigration that could confound results. Our objective was to apply graph theory to a dataset spanning a 9-year period (2002–2010) to identify specific regions in the NLP that function as source and sink areas. In addition, to demonstrate the utility of our approach, we present a case study using remotely sensed and biological data to examine the importance of landscape and ecological features on directional dispersal and show how graphs can be used to provide insight into the mechanisms underlying source–sink dynamics.

## 2. Material and methods

### (a) Study area

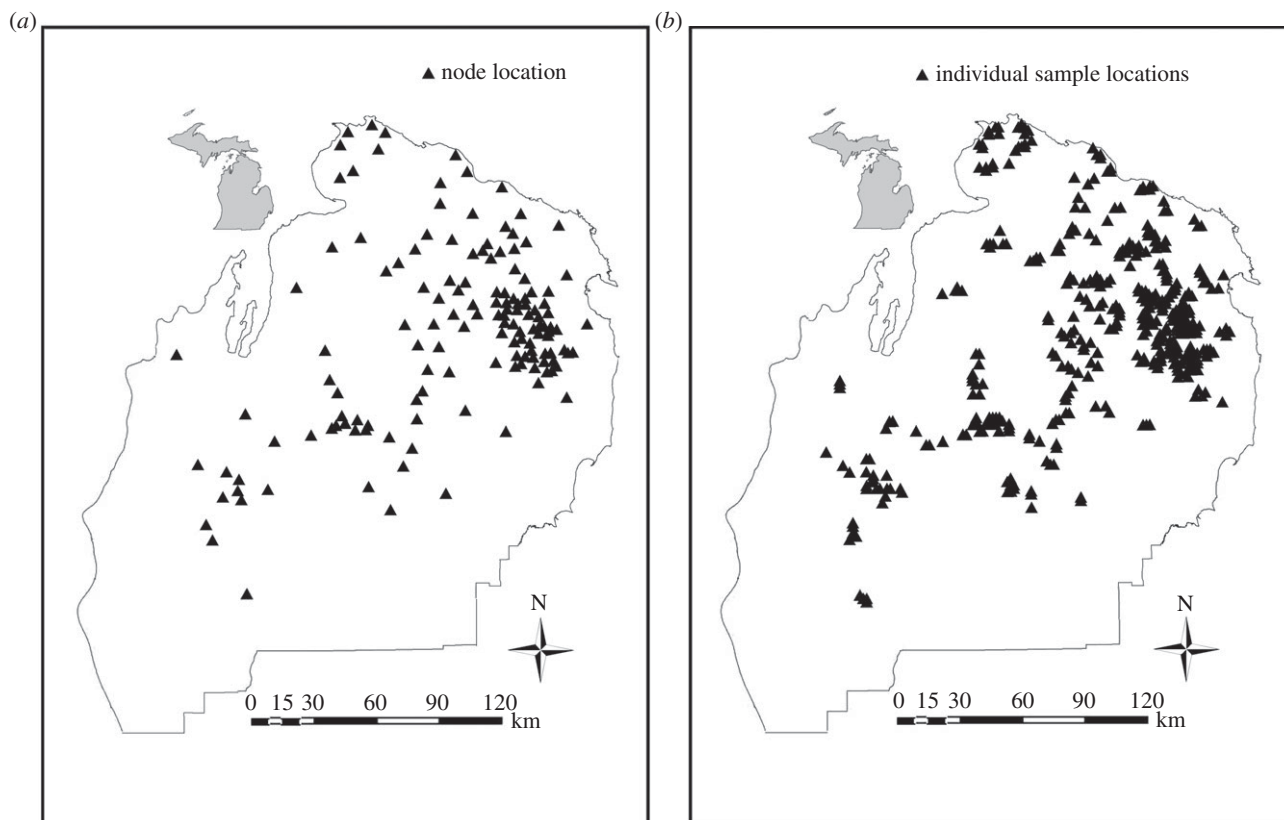
Michigan's NLP is a 47 739 km<sup>2</sup> (figure 1) fragmented mosaic of variable land cover and land use types including development, agriculture, upland non-forested openings, northern hardwood and mixed hardwood, oak, aspen, pine, forested wetland, and non-forested wetland (electronic supplementary material, figure S1) [25]. Teeth were collected from harvested black bears registered at hunter check stations during autumn harvest (September–October) of 2002 ( $n = 263$ ), 2006 ( $n = 385$ ) and 2010 ( $n = 336$ ). Locations of bear harvest samples were recorded as township, range and sections, which were then georeferenced to the section (area of a section = 2.6 km<sup>2</sup>) centroid and converted into UTM coordinates.

### (b) Laboratory analysis

We extracted DNA from bear teeth using Qiagen DNEasy Tissue Kits following manufacturer protocols (Qiagen Inc., Valencia, CA). DNA was quantified using a Nanodrop spectrophotometer (Thermo Scientific, Waltham, MA) and diluted to a 20 ng  $\mu\text{l}^{-1}$  working concentration. We amplified 12 microsatellite loci using polymerase chain reaction (PCR), including G10X, G10L, G10D, G10M, G10B [28], UarMU50 and UarMU59 [29], UT29, UT35, UT38 [30], ABB1 and ABB4 [31], using methods described by Moore *et al.* [32]. In total, 10% of samples were randomly selected and genotyped twice to provide a genotyping error rate of less than 2%. Using MICRO-CHECKER [33] and GENALEX v. 6.0, [34], no loci were found to deviate significantly from Hardy–Weinberg and linkage equilibrium, so all 12 were retained for further analyses (electronic supplementary material, table S1).

### (c) Defining the nodes

Graphs are composed of nodes connected by edges, and both can be assigned weights. When applied in a landscape context, nodes are traditionally defined as the spatial centroid of a patch or population, and are generally weighted by attributes representing measures of quality (e.g. patch size or habitat suitability) [13]. As it is not feasible to assign node weights representing local landscape or ecological characteristics to a location representing a single individual, we defined nodes as areas that were consistently occupied in both space and time. Our sample includes individuals collected from 2002, 2006 and 2010, representing three generations in this population (average age of reproductive females = 3 [35]). We used the harvest locations from each year to generate polygon layers using Voronoi tessellations. Briefly, Voronoi tessellations are polygons created around each point such that the borders of each polygon are equidistant from their two nearest points. Annual polygon layers were then overlaid and areas of consistent occupancy were identified where polygons from 2002, 2006 and 2010 overlapped and contained at least one bear from each of those years. We used STAMP (spatial-temporal analysis of



**Figure 1.** Study area in the northern lower peninsula (NLP) of Michigan showing (a) node locations ( $n = 141$ ) and (b) sample locations of black bear harvest samples collected during 2002, 2006 and 2010 ( $n = 569$ ).

moving polygons [36]) in ArcGIS v. 10.0 to identify overlap areas. Nodes were thus defined as areas of consistent polygon overlap (i.e. occupancy) across the three time points (electronic supplementary material, figure S2). Node locations were defined as the centroids of overlapping polygons that contained at least one bear from 2002, 2006 and 2010.

#### (d) Graph construction

We generated a genetic graph and a set of ecological graphs to explore genetic, habitat and demographic connectivity of the NLP black bear population:

- (1) *Genetic modelling.* We calculated pairwise relatedness among individuals using a maximum-likelihood estimator of relatedness ( $K$  [37]), in the program ML-RELATE [38]. We constructed a genetic ‘saturated graph’ in which each node was connected to every other node. We assigned node weights as the mean relatedness among individuals within each node, and edge weights as the mean pairwise relatedness between individuals at node  $i$  and node  $j$ . We tested for effects of sample size on mean relatedness at the nodes (node weight) and found no significant association (linear regression;  $R^2 = 0.007$ ,  $p = 0.50$ ).
  - (2) *Ecological modelling.* We created ecological graphs to test associations between factors that affect node importance (strength) and overall graph connectivity. Edges were calculated using Euclidean or least-cost distance (based on habitat permeability) from node  $i$  to node  $j$ . We assigned node weights based on habitat suitability or local harvest density (a proxy for bear population density). We performed Spearman’s rank correlation analyses between the genetic model and our set of ecological candidate models (table 1) in R (R Development Core Team, 2012).
- (a) *Weighting nodes.* To obtain weights for each node based on measures of habitat suitability, we reclassified the 2006

National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (C-CAP) Land Cover dataset (resolution = 30 m) into eight land cover classifications and ranked them based on bear habitat suitability (1–100, least to most suitable; electronic supplementary material, table S2) including: mixed deciduous forest (MF, 100), forested wetland (FW, 100), evergreen forest (EF, 90), non-forested upland (NFU, 80), agriculture (AG, 50), non-forested wetland (NFW, 1), developed (DEV, 1) and open water (NA) using a habitat suitability model developed independently for Michigan’s NLP black bears [25]. In addition, we reclassified land cover data into a coarser habitat suitability classification as either bear habitat (MF, FW; 100) or non-habitat (EF, NFU, AG, NFW, DEV; 1).

We used localized harvest density, based on bear harvest locations, as a proxy for local bear population density, as in Moore *et al.* [32,39]. To estimate local harvest density we used annual harvest locations from 2002–2010, generated kernel density grids in ArcGIS v. 10.0 and reclassified them into a range of 1–10 (low to high as previously defined [32]). We then created a harvest grid by calculating the median values over the nine annual harvest grids (electronic supplementary material, figure S3). We extracted a local harvest density and land cover value based on majority of grid cells within the nodes.

- (b) *Weighting edges.* Edges for ecological graphs were weighted by either Euclidean distance or least-cost distance (one path of least resistance through a resistance surface, reflecting permeability through habitat types or boundaries) in ArcGIS v. 10.0. Least-cost distance was calculated using three resistance surfaces to model habitat permeability (table 1). The 2006 CCAP dataset was reclassified into eight cover types reflecting the ecological model [25] and two coarse cover types (bear habitat

**Table 1.** Spearman's rank correlations between genetic and ecological measures of node importance (source strength). The three graph metrics compared are net flux ( $F_{\text{out}} - F_{\text{in}}$ ), area-weighted flux (AWF) and probability of connectivity (PC). 'High' indicates maximum magnitude weights (e.g. 1–100) and 'low' indicates minimum magnitude weights (e.g. 1–5) used to test the sensitivity of analyses to the magnitude of cost values. All correlations were significant ( $p < 0.05$ ).

ecological model		graph metrics		
node weight	edge weight	$F_{\text{out}} - F_{\text{in}}$	AWF	PC
habitat suitability (high)	habitat suitability (high)	0.1865	0.2040	0.2094
habitat suitability (low)	habitat suitability (low)	0.1947	0.2160	0.2050
habitat suitability (high)	road (high)	0.1836	0.2270	0.2160
habitat suitability (low)	road (low)	0.1809	0.2150	0.2040
habitat suitability (high)	distance only	0.1799	0.2270	0.1980
habitat/non-habitat (high)	habitat/non-habitat (high)	0.1712	0.2030	0.1985
habitat/non-habitat (low)	habitat/non-habitat (low)	0.1690	0.2100	0.2081
habitat/non-habitat (high)	road (high)	0.1681	0.2120	0.1970
habitat/non-habitat (low)	road (low)	0.1581	0.2040	0.2030
habitat/non-habitat (high)	distance only	0.1490	0.2100	0.1970
density	habitat suitability (high)	0.2021	0.2630	0.2590
density	habitat suitability (low)	0.2032	0.2780	0.2750
density	habitat/non-habitat (high)	0.2020	0.2740	0.2700
density	habitat/non-habitat (low)	0.2020	0.2820	0.2770
density	road (high)	0.2034	0.2800	0.2760
density	road (low)	0.2020	0.2830	0.2790
density	distance only	0.2021	0.2820	0.2780

versus non-habitat; table 1). In addition, we obtained and reclassified a road-only resistance surface from the Michigan Geographic Data Library (Center for Geographic Information; electronic supplementary material, table S2). Pairwise least-cost distances were calculated using the landscape genetic toolbox in ARCGIS [40]. Least-cost analysis and node weights may be sensitive to relative cost values assigned to land cover types [41] and could potentially influence our results. Thus, we performed a sensitivity analysis on the range of relative cost values between resistance surface land cover types by comparing least-cost distances between rasters weighted using large differences (i.e. maximum weights = 1, 10, 20, 50, 100) and small differences (minimum weights = 1, 2, 3, 4, 5) in the relative cost values for land cover types (table 1).

### (e) Node importance and source–sink analysis

Graph metrics were calculated to define node importance to identify source areas. We calculated dispersal probability ( $p_{ij}$ ) for each pair of nodes,

$$p_{ij} = \exp(k \times d_{ij}), \quad (2.1)$$

where  $k$  is a decay coefficient calculated from a tail dispersal distance (an arbitrarily selected point on the flat tail of a negative-exponential dispersal-distance function) and  $d_{ij}$  is the distance from node  $i$  to  $j$  [42]. Area-weighted flux (AWF) was estimated using the previous metric (dispersal probability) multiplied by the quality area (i.e. genetic graph,  $q_i$  = mean relatedness; ecological graph,  $q_i$  = density or habitat quality) of the nodes [43],

$$\text{AWF} = q_i \times p_{ij}. \quad (2.2)$$

Probability of connectivity (PC) is a graph metric that combines the attributes of the nodes with the maximum product probability of all the possible paths between every pair of nodes,

$$\text{PC} = \frac{\sum_{i=1}^n \sum_{j=1}^n a_i a_j p_{ij}}{A_L^2}, \quad (2.3)$$

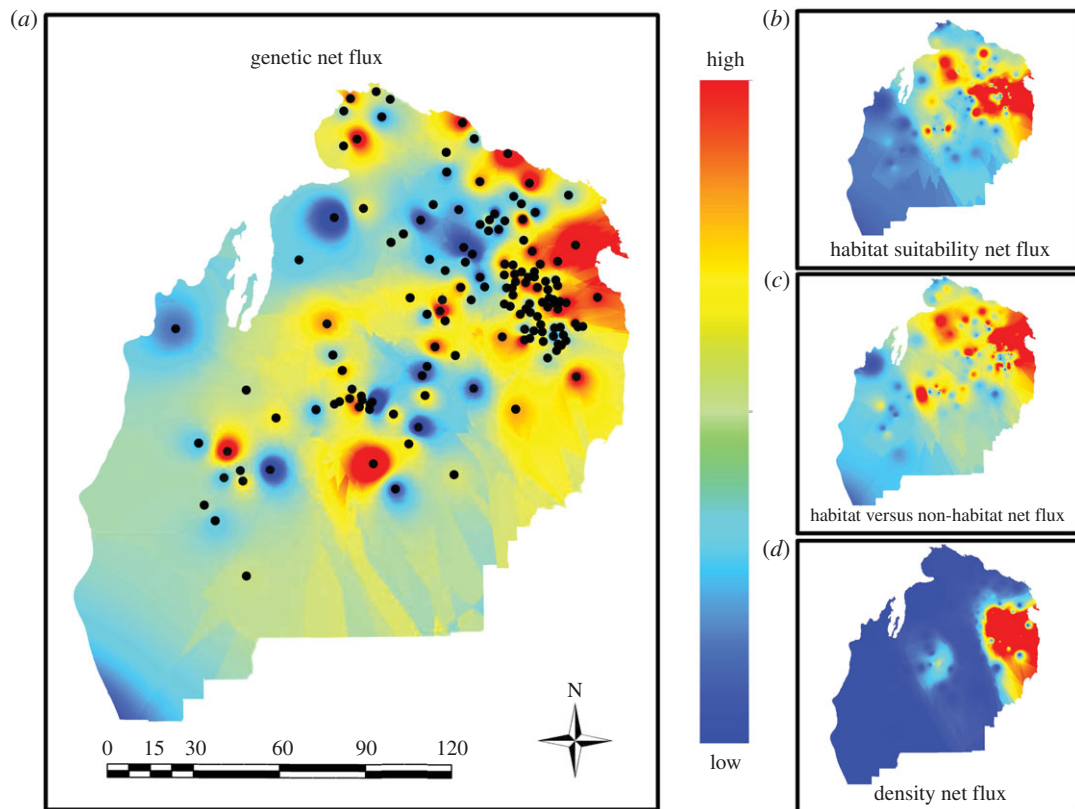
where  $A_L$  is the total quality measure of the entire graph, and  $a_i$  and  $a_j$  are the qualities of node  $i$  and  $j$ , respectively [44]. AWF and PC were calculated using program CONEFOR v. 2.6 [45]. Influx (flux entering a node) and outflux (flux exiting a node) were calculated to account for asymmetric dispersal as there are two distinct edges for each pair of nodes. Edges were based on distances so that  $p_{ij} = p_{ji}$ . However, influx and outflux at a particular node may differ, as flux also incorporates quality area  $q_i$  of the donor node  $i$  (i.e.  $f_{ij} = q_i \times p_{ij}$  versus  $f_{ji} = q_j \times p_{ij}$ ). Total flux out of node  $i$  and total flux into node  $i$  were calculated as

$$F_{\text{out } i} = \sum f_{ji} \quad (2.4)$$

and

$$F_{\text{in } i} = \sum f_{ij}. \quad (2.5)$$

To determine the magnitude and direction of the net flux associated with each node, we calculated the flux difference ( $F_{\text{out } i} - F_{\text{in } i}$ ) [43]. High outflux indicates a large number of emigrants and high influx indicates a large number of immigrants. Thus, source nodes were characterized by high total outflux (positive net flux,  $F_{\text{out } i} > F_{\text{in } i}$ ) and sink nodes were characterized by high influx and low outflux (negative net flux,  $F_{\text{out } i} < F_{\text{in } i}$ ). To visualize the spatial configuration of source and sink nodes, we generated interpolated surfaces of node importance (source strength) using inverse distance weighting (IDW).



**Figure 2.** Interpolated surfaces of black bear node importance (source strength) based on net flux at the nodes across the NLP study area using inverse distance weighting (IDW). Interpolated surfaces of (a) the genetic model, (b) Carter *et al.* [25] habitat suitability model, (c) a habitat/non-habitat model and (d) a bear density model are shown for comparison. Warm colours represent sources and cool colours represent sinks. Circles are the node locations.

### (f) Resampling simulation

Harvest samples are opportunistically collected, and therefore samples may be predisposed to erroneous inference due to uneven sampling commonly encountered in empirical datasets. To test whether our approach is robust to uneven sample density and dispersion, we performed a series of simulations by creating 10 randomly resampled datasets under two different sampling scenarios. The two scenarios differed with respect to the degree with which sampling was correlated with population density. The first scenario was a stratified resampling approach where sample density and dispersion were weakly correlated or uncorrelated with population density. The second scenario was a density-biased sampling where sample density and dispersion were correlated with population density. The second scenario reflects our original dataset. For stratified sampling, we randomly sampled 125 individuals without replacement (i.e. individuals could not be selected more than once per iteration) from the 2002, 2006 and 2010 empirical genetic datasets, and enforced a 7 km minimum distance among sampling locations. This parameter constrained the density and distribution of samples such that locations were at least 7 km from each other. We chose a minimum distance of 7 km (i) to ensure a more even sample density and dispersion and (ii) to capture the scale of the spatial processes known to drive genetic patterns (e.g. black bears exhibit female philopatry and median female bear dispersal distance in our study population is approximately 7 km [35]). For density-biased resampling, we performed weighted randomly sampling without replacement (125 individuals from the 2002, 2006, and 2010 genotyped datasets) so the probability of each sample being selected was influenced by bear density (i.e. harvest samples collected in areas of higher bear density had a greater probability of being selected than samples collected in areas of lower sample density). Sample weights were assigned by overlaying the empirical datasets on the median density grid (electronic supplementary

material, figure S3) and assigning the corresponding density values of the intersecting raster cell. All analyses were performed using the *sampling* package in R v. 3.0.1 (R Development Core Team, 2013) and ArcGIS v. 10.0. For each resampling iteration for both scenarios, we then used previously described methods to generate genetic graphs, net flux values and interpolated surfaces.

## 3. Results

Based on our saturated genetic graph using 141 nodes (areas of consistent occupancy), figure 2 presents the juxtaposition of inferred source and sink nodes over the study area. Of the 141 nodes in the NLP, 62 are considered source nodes (positive net flux), with 16 nodes having net flux more than 0.7. By contrast, 79 nodes were classified as sinks (negative net flux). We estimated two additional graph metrics that measure node importance (i.e. source strength): (i) AWF and (ii) PC. AWF and PC were highly correlated with net flux (Spearman's rank correlations  $\rho > 0.90$ ,  $p < 0.05$ ).

Interpolated surfaces of net flux for each sampling iteration indicated that areas of high and low net flux are consistent among original (full) and simulated (resampled) datasets (figure 2; electronic supplementary material, figure S4a–d). For all datasets, we observed clusters of source nodes in the northern, east-central and south-central portions of the study area. We did observe fine-scale spatial variation among simulation runs (i.e. specific locations of source and sink nodes differed slightly), which was probably an effect of random sampling causing actual node locations to differ between iterations (electronic supplementary material, figure S4c–d). Nonetheless, our results

suggest that broad-scale source–sink patterns will be detected irrespective of whether sample dispersion is correlated with population density.

We constructed additional saturated graphs based on the same 141 nodes using purely ecological data (only habitat or local harvest density). Correlation analyses of node metrics derived from genetic and ecological graphs revealed four major findings. First, correlations between graph metrics (net flux, AWF and PC) and habitat/local harvest abundance graph metrics were all significant ( $\rho = 0.149\text{--}0.283$ ,  $p < 0.05$ ; table 1). Second, relative to associations with habitat suitability, genetic graph metrics were most correlated with local bear harvest density (net flux, mean  $\rho = 0.202$ ; AWF, mean  $\rho = 0.277$ ; PC, mean  $\rho = 0.273$ ,  $p < 0.05$ ). Third, ecological models where nodes were weighted by habitat suitability had comparable correlation values irrespective of how finely or coarsely the land cover was defined (habitat suitability model: net flux, mean  $\rho = 0.185$ ; AWF, mean  $\rho = 0.218$ ; PC, mean  $\rho = 0.207$ ; habitat versus non-habitat: net flux, mean  $\rho = 0.163$ ; AWF, mean  $\rho = 0.210$ ; PC, mean  $\rho = 0.201$ ; table 1). Lastly, because least-cost distances can be sensitive to the magnitude of cost values, we performed a sensitivity analysis of cost values assigned to land cover and found graph metrics were not sensitive to the range of land cover cost values used (table 1).

#### 4. Discussion

By combining genetic and ecological graphs, we have identified source and sink areas using individual-based genetic data in a landscape genetic framework. Our methodological challenge was to find a way to apply genetic graphs to a continuously distributed population exhibiting clinal IBD. If populations are patchily distributed and/or strongly genetically structured, allowing for *a priori* delineation of populations or patches, genetic weights can easily be assigned by calculating the mean genetic similarity (e.g. relatedness) over the entire population or patch [2,16]. However, for continuously distributed species that cannot be unambiguously placed into defined populations or patches, such as black bears, an alternative, individual-based method of weighting nodes is necessary. We overcame this methodological obstacle by using point pattern and spatial analyses to identify areas of consistent spatial occupancy over time and assigned node weights based on average pairwise relatedness of individuals within the nodes.

Genetic graphs can be applied to detect source–sink dynamics in other systems where populations are not patchily distributed or strongly genetically structured into discrete populations. For example, our method may find utility in wide-ranging and elusive species where source–sink dynamics has gone undetected (e.g. fur-bearing species, carnivores, marine mammals [46–48]). Undetected source–sink dynamics has potentially negative consequences, particularly for harvested species. For example, if the management goal is to reduce abundance in a specific sink area through harvest, dispersal from surrounding source areas will confound efforts. Alternatively, a source area with abundant resources and subsequent high reproduction may become an ‘attractive sink’ if harvest pressure is high [48].

Our analyses revealed patterns of asymmetric connectivity in NLP black bears based on areas of high and low net flux. Findings are consistent with previous local spatial

autocorrelation analyses. Draheim *et al.* [49] found evidence of fine-scale spatial genetic structure in NLP black bears but no strong population structuring across the region. We identified 62 source areas including 16 nodes we consider to be important source areas (net flux  $> 0.7$ ) in the NLP. Interpolated surfaces of net flux values (figure 2) revealed important source areas were patchily distributed with clusters of adjacent high net flux nodes occurring on private and forested public lands in the northwestern, northern-central and south-central portions of the study area.

Figure 2 is interpolated from point data ( $n = 141$  nodes), and observed patterns are dependent on the spatial dispersion of nodes (figure 1). Therefore, interpolated maps should be interpreted as displaying regional patterns and not the exact locations of source and sink nodes. Net flux estimates along the study area boundary should be interpreted with caution as interpolations may be prone to edge effects [50]. Harvest sample collection is often opportunistic and location may reflect hunter behaviour. Accordingly, because bear occupancy is based on location of harvest samples, the spatial structure of sources and sinks may not include all relevant locations where bears are present. However, Draheim *et al.* [51] found that black bear sample density and distribution is consistent when samples are opportunistically collected via harvest or collected via systematic non-invasive hair snares. Indeed, for NLP black bears, opportunistic and systematic sampling capture spatial genetic heterogeneity at the same scale and spatial extent [51].

The spatial dispersion of source and sink areas do not appear to be unduly influenced by sampling. Through resampling simulations, we detected asymmetric connectivity regardless of whether the sample density and dispersion was correlated with population density. Interpolated maps based on net flux values show the same regional patterns of high and low net flux as the original dataset, whether we used stratified or density-biased sampling scenarios (electronic supplementary material, figure S4*a,b*). We note, however, that close visual inspection of all interpolated maps does show some variation in fine-scale spatial patterns among resampling iterations and among sampling regimes (electronic supplementary material, figure S4*c,d*). Smaller sample sizes and reduced power are probably impacting our resolution and ability to consistently discern fine-scale patterns. Additionally, because spatial dispersion of the nodes is determined by the underlying sample locations, subsampling will cause the locations of the node centroids to differ slightly between iterations based on the samples that were selected to define the nodes.

We found significant correlations between graph metrics derived from genetic data and ecological models based on habitat quality and local harvest density. Our correlation values were moderate. Low correlations between genetic and habitat metrics are not surprising as much of the NLP is classified as high-quality habitat. Indeed, as shown in table 1, how habitat was classified and subsequently weighted did not appreciably affect model outcomes. In addition, there is a degree of uncertainty in how representative land cover reflects resource availability [52].

Local harvest density at the nodes better explained source strength than did habitat quality. We predicted source areas identified using genetic data would be characterized by areas of high local harvest abundance, which can be used as a surrogate for population density [53,54]. As an indirect estimate, we acknowledge that local harvest density may

not be a strong index of bear density and could potentially introduce noise into the correlational relationship between the genetic and density graph metrics. However, associations between local harvest density and bear density are predicted based on several factors. First, as mentioned above, density of bear samples collected via harvest was concordant with the density of samples collected from a study using systematically placed hair snares and the geographical distribution of harvest locations shows little annual variation [51]. Second, the NLP population undergoes high annual harvest in which 13–29% of the population is harvested each year. At this level of harvest, local harvest density would probably vary annually if it were not correlated with bear density. Third, there are no large areas that are not accessible to hunters in a motorized vehicle, thus there is no refuge for black bears to escape hunter pressure (electronic supplementary material, figure S3). Fourth, land ownership does not limit hunter access as approximately 48–62% of hunters used private land during harvest [55]. Finally, local harvest density has repeatedly been found to be associated with other features of the NLP population, including natal dispersal distance [32] and mate selection [39], based on pedigree analysis.

Our approach is centred on genetic data; therefore, our level of inference is limited to spatial patterns of asymmetric gene flow. Indeed, rigorous identification of source and sink areas requires a great deal of detailed demographic information on population survival rates, fecundity, immigration and emigration [3,6,56]. However, such detailed information is difficult to acquire for an elusive, highly mobile animal such as the black bear. Dispersal studies that use traditional methods (i.e. capture–mark–recapture or radio-telemetry) are expensive and labour-intensive [57], and may lead to underestimates due to small sample size and limited geographical scope. For conservation and management, one compelling feature of a graph framework is the ability to identify nodes that are important for landscape-scale connectivity, using graph metrics or node removal. Endangered species or species of conservation concern areas, which are critical for maintaining connectivity, can be identified and given conservation preference (e.g. [58]). Alternatively, as a form of biological control to prevent spread of pests [59], invasive species [60] or pathogens [61], habitat patches could be removed when connectivity is detrimental. The NLP black bear population is subjected to two factors that could impact connectivity:

intensive harvest and landscape modification. Identification of source areas could guide managers in determining boundaries for bear management units, in setting regional harvest quotas, and in identifying areas where landscape change may have a disproportionate effect on black bear connectivity.

## 5. Conclusion

Graphs are simplified portrayals of natural systems that we have shown can be powerful tools for assessing and detecting source–sink dynamics [42,43]. Our method does not estimate migration rates. Rather, we derived graph metrics based on pairwise relatedness within and among nodes to evaluate plausible spatially explicit hypotheses about degree and directionality of connectivity among nodes (source–sink dynamics), and the ecological and environmental factors that affect connectivity. Genetic graphs provide a viable alternative method for detecting source–sink dynamics in continuously distributed species. Because genetic graphs provide a flexible framework for understanding connectivity, they could be widely integrated into landscape genetics research and conservation planning at multiple spatial scales. While genetic graphs may be broadly applicable across a range of taxa, the graph models need to be parametrized based on the specific life-history characteristics of the species or population of interest.

**Ethics.** The research presented here was approved by the Institutional Animal Care and Use Committee of Michigan State University.

**Data accessibility.** Data used in this study, including bear microsatellite genotypes and harvest locations, are available for download on Dryad (<http://dx.doi.org/10.5061/dryad.c61q0>).

**Authors' contributions.** The authors jointly conceived of the ideas presented in this study. Molecular work was conducted by H.M.D and J.A.M. H.M.D. conducted all genetic and graph analyses. All authors contributed to the writing of the manuscript.

**Competing interests.** We declare we have no competing interests.

**Funding.** Support for this project was provided by the Michigan Department of Natural Resources through the Wildlife and Sportfish Restoration Program F11AF00640 and by the Department of Fisheries and Wildlife at Michigan State University.

**Acknowledgements.** We are grateful to J. Fierke, K. Filcek, B. Dreher and S. Libants for assisting in data generation. We also thank M. Fortin, J. Messina, B. Epperson, K. Holekamp, J. Kanefsky and J. McGuire for insightful discussion and helpful comments on the analysis and manuscript. All samples used in analyses were provided by Michigan bear hunters and collected by the many MDNR staff at registration stations.

## References

- Knowlton JL, Graham CH. 2010 Using behavioral landscape ecology to predict species' responses to land-use and climate change. *Biol. Conserv.* **143**, 1342–1354. (doi:10.1016/j.biocon.2010.03.011)
- Clobert J, Le Galliard JF, Cote J, Meylan S, Massot M. 2009 Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecol. Lett.* **12**, 197–209. (doi:10.1111/j.1461-0248.2008.01267.x)
- Pulliam HR. 1988 Sources, sinks, and population regulation. *Am. Nat.* **132**, 652–661. (doi:10.1086/284880)
- Minor ES, Urban DL. 2008 A graph-theory framework for evaluating landscape connectivity and conservation planning. *Conserv. Biol.* **22**, 297–307. (doi:10.1111/j.1523-1739.2007.00871.x)
- Boughton DA. 1999 Empirical evidence for complex source–sink dynamics with alternative states in a butterfly metapopulation. *Ecology* **80**, 2727–2739. (doi:10.1890/0012-9658(1999)080[2727:EEFCSS]2.0.CO;2)
- Dias PC. 1996 Sources and sinks in population biology. *Trends Ecol. Evol.* **11**, 326–330. (doi:10.1016/0169-5347(96)10037-9)
- Gaggiotti OE. 1996 Population genetic models of source–sink metapopulations. *Theor. Popul. Biol.* **50**, 178–208. (doi:10.1006/tpbi.1996.0028)
- Anderson JT, Geber MA. 2010 Demographic source–sink dynamics restrict local adaptation in Elliott's blueberry (*Vaccinium elliotii*). *Evolution* **64**, 370–384. (doi:10.1111/j.1558-5646.2009.00825.x)
- Kawecki T. 1995 Demography of source–sink populations and the evolution of ecological niches. *Evol. Ecol.* **9**, 38–44. (doi:10.1007/bf01237695)
- Kawecki TJ. 2008 Adaptation to marginal habitats. *Annu. Rev. Ecol. Evol. Syst.* **39**, 321–342. (doi:10.1146/annurev.ecolsys.38.091206.095622)
- Lenormand T. 2002 Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**, 183–189. (doi:10.1016/S0169-5347(02)02497-7)

12. Peery MZ, Beissinger SR, House RF, Berube M, Hall LA, Sellas A, Palsboll PJ. 2008 Characterizing source–sink dynamics with genetic parentage assignments. *Ecology* **89**, 2746–2759. (doi:10.1890/07-2026.1)
13. Dale MRT, Fortin M-J. 2010 From graphs to spatial graphs. *Annu. Rev. Ecol. Syst.* **41**, 21–38. (doi:10.1146/annurev-ecolsys-102209-144718)
14. Peterman W, Rittenhouse TG, Earl J, Semlitsch R. 2013 Demographic network and multi-season occupancy modeling of *Rana sylvatica* reveal spatial and temporal patterns of population connectivity and persistence. *Landscape Ecol.* **28**, 1601–1613. (doi:10.1007/s10980-013-9906-9)
15. Rayfield B, Fortin M-J, Fall A. 2011 Connectivity for conservation: a framework to classify network measures. *Ecology* **92**, 847–858. (doi:10.1890/09-2190.1)
16. Murphy MA, Dezzani R, Pilliod DS, Storer A. 2010 Landscape genetics of high mountain frog metapopulations. *Mol. Ecol.* **19**, 3634–3649. (doi:10.1111/j.1365-294X.2010.04723.x)
17. 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. (doi:10.1086/506976)
18. Rousset 2000 Genetic differentiation between individuals. *J. Evol. Biol.* **13**, 58–62. (doi:10.1046/j.1420-9101.2000.00137.x)
19. Queller DC, Goodnight KF. 1989 Estimating relatedness using genetic markers. *Evolution* **43**, 258–275. (doi:10.2307/2409206)
20. Garroway 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. (doi:10.1111/j.1365-294X.2011.05243.x)
21. Wright S. 1943 Isolation by distance. *Genetics* **28**, 114–138.
22. Manel S, Schwartz MK, Luikart G, Taberlet P. 2003 Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* **18**, 189–197. (doi:10.1016/S0169-5347(03)00008-9)
23. Keller D, Holderegger R, van Strien MJ, Bolliger J. 2014 How to make landscape genetics beneficial for conservation management? *Conserv. Genet.* **16**, 503–512. (doi:10.1007/s10592-014-0684-y)
24. van Strien MJ, Keller D, Holderegger R, Ghazoul J, Kienast F, Bolliger J. 2014 Landscape genetics as a tool for conservation planning: predicting the effects of landscape change on gene flow. *Ecol. Appl.* **24**, 327–339. (doi:10.1890/13-0442.1)
25. Carter NH, Brown DG, Etter DR, Visser LG. 2010 American black bear habitat selection in northern Lower Peninsula, Michigan, USA, using discrete-choice modeling. *Ursus* **21**, 57–71. (doi:10.2192/09GR011.1)
26. Costello CM. 2010 Estimates of dispersal and home-range fidelity in American black bears. *J. Mammal* **91**, 116–121. (doi:10.1644/09-Mamm-a-015r1.1)
27. Horner MA, Powell RA. 1990 Internal structure of home ranges of black bears and analyses of home-range overlap. *J. Mammal* **71**, 402–410. (doi:10.2307/1381953)
28. Paetkau D, Calvert W, Stirling I, Strobeck C. 1995 Microsatellite analysis of population-structure in Canadian polar bears. *Mol. Ecol.* **4**, 347–354. (doi:10.1111/j.1365-294X.1995.tb00227.x)
29. Taberlet P, Camarra JJ, Griffin S, Uhres E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J. 1997 Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Mol. Ecol.* **6**, 869–876. (doi:10.1046/j.1365-294X.1997.00251.x)
30. Shih CH, Huang CC, Li SH, Hwang MH, Lee LL. 2009 Ten novel tetranucleotide microsatellite DNA markers from Asiatic black bear, *Ursus thibetanus*. *Conserv. Genet.* **10**, 1845–1847. (doi:10.1007/s10592-009-9830-3)
31. Wu H, Zhang SN, Wei FW. 2010 Twelve novel polymorphic microsatellite loci developed from the Asiatic black bear (*Ursus thibetanus*). *Conserv. Genet.* **11**, 1215–1217. (doi:10.1007/s10592-009-9922-0)
32. Moore JA, Draheim HM, Etter D, Winterstein S, Scribner KT. 2014 Application of large-scale parentage analysis for investigating natal dispersal in highly vagile vertebrates: a case study of american black bears (*Ursus americanus*). *PLoS ONE* **9**, e91168. (doi:10.1371/journal.pone.0091168)
33. 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. (doi:10.1111/j.1471-8286.2004.00684.x)
34. Peakall R, Smouse PE. 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295. (doi:10.1111/j.1471-8286.2005.01155.x)
35. Etter DR, Visser LG, Schumacher CM, Carlson E, Reis T, Rabe D. 2002 *Black bear population management techniques*. Federal Aid in Wildlife Restoration Project W-127-R-20. Lansing, MI: Michigan Department of Natural Resources.
36. Robertson C, Nelson T, Boots B, Wulder M. 2007 STAMP: spatial–temporal analysis of moving polygons. *J. Geogr. Syst.* **9**, 207–227. (doi:10.1007/s10109-007-0044-2)
37. Wagner AP, Creel S, Kalinowski ST. 2006 Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity* **97**, 336–345. (doi:10.1038/sj.hdy.6800865)
38. Kalinowski ST, Wagner AP, Taper ML. 2006 ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Mol. Ecol. Notes* **6**, 576–579. (doi:10.1111/j.1471-8286.2006.01256.x)
39. Moore JA, Xu R, Frank K, Draheim H, Scribner KT. 2015 Social network analysis of mating patterns in American black bears (*Ursus americanus*). *Mol. Ecol.* **24**, 4010–4022. (doi:10.1111/mec.13290)
40. Etherington TR. 2011 Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. *Methods Ecol. Evol.* **2**, 52–55. (doi:10.1111/j.2041-210X.2010.00048.x)
41. Rayfield B, Fortin M-J, Fall A. 2010 The sensitivity of least-cost habitat graphs to relative cost surface values. *Landscape Ecol.* **25**, 519–532. (doi:10.1007/s10980-009-9436-7)
42. Urban D, Keitt T. 2001 Landscape connectivity: a graph-theoretic perspective. *Ecology* **82**, 1205–1218. (doi:10.2307/2679983)
43. Minor ES, Urban DL. 2007 Graph theory as a proxy for spatially explicit population models in conservation planning. *Ecol. Appl.* **17**, 1771–1782. (doi:10.1890/06-1073.1)
44. Saura S, Pascual-Hortal L. 2007 A new habitat availability index to integrate connectivity in landscape conservation planning: comparison with existing indices and application to a case study. *Landscape Urban Plan* **83**, 91–103. (doi:10.1016/j.landurbplan.2007.03.005)
45. Saura S, Torne J. 2009 Conefor Sensinode 2.2: a software package for quantifying the importance of habitat patches for landscape connectivity. *Environ. Model. Softw.* **24**, 135–139. (doi:10.1016/j.envsoft.2008.05.005)
46. Hoffman JI, Matson CW, Amos W, Loughlin TR, Bickham JW. 2006 Deep genetic subdivision within a continuously distributed and highly vagile marine mammal, the Steller's sea lion (*Eumetopias jubatus*). *Mol. Ecol.* **15**, 2821–2832. (doi:10.1111/j.1365-294X.2006.02991.x)
47. Tammeleht E, Remm J, Korsten M, Davison J, Tumanov I, Saveljev A, Mannil P, Kojala I, Saarma U. 2010 Genetic structure in large, continuous mammal populations: the example of brown bears in northwestern Eurasia. *Mol. Ecol.* **19**, 5359–5370. (doi:10.1111/j.1365-294X.2010.04885.x)
48. Robinson HS, Wielgus RB, Cooley HS, Cooley SW. 2008 Sink populations in carnivore management: cougar demography and immigration in a hunted population. *Ecol. Appl.* **18**, 1028–1037. (doi:10.1890/07-0352.1)
49. Draheim HM. 2015 Landscape genetics of black bears (*Ursus Americanus*) in the Northern Lower Peninsula (NLP) of Michigan, USA. Dissertation, Michigan State University, East Lansing, MI.
50. Okabe A, Boots B, Sugihara KSN. 2000 *Spatial tessellations: concepts and applications of voronoi diagrams*, 2nd edn, pp. 411–452. Chichester, UK: Wiley.
51. Draheim HM, Lopez V, Winterstein SR, Etter DR, Scribner KT. 2015 Effects of spatial and temporal sampling scale on sample dispersion and spatial genetic structure. *Ursus* **26**, 143–156. (doi:10.2192/URSUS-D-15-00011.1)
52. Booth DT, Cox SE, Meikle T, Zuuring HR. 2008 Ground-cover measurements: assessing correlation among aerial and ground-based methods. *Environ. Manage.* **42**, 1091–1100. (doi:10.1007/s00267-008-9110-x)
53. Royama T. 1992 *Analytical population dynamics*. London, UK: Chapman and Hall.
54. Cattadori IM, Haydon DT, Thirgood SJ, Hudson PJ. 2003 Are indirect measures of abundance a useful index of population density? The case of red grouse harvesting. *Oikos* **100**, 439–446. (doi:10.1034/j.1600-0706.2003.12072.x)
55. Frawley BJ. 2010 *2010 Michigan black bear hunter survey*, 29. Lansing, MI: MDNR, Wildlife Division.



56. Lowe WH, Allendorf FW. 2010 What can genetics tell us about population connectivity? (*Mol. Ecol.* **19**, 3038–3051). *Mol. Ecol.* **19**, 5320. (doi:10.1111/j.1365-294X.2010.04878.x)
57. Solberg KH, Bellemain E, Drageset OM, Taberlet P, Swenson JE. 2006 An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*) population size. *Biol. Conserv.* **128**, 158–168. (doi:10.1016/j.biocon.2005.09.025)
58. Pascual-Hortal L, Saura S. 2006 Comparison and development of new graph-based landscape connectivity indices: towards the prioritization of habitat patches and corridors for conservation. *Landscape Ecol.* **21**, 959–967. (doi:10.1007/s10980-006-0013-z)
59. Koh I, Rowe HI, Holland JD. 2013 Graph and circuit theory connectivity models of conservation biological control agents. *Ecol. Appl.* **23**, 1554–1573. (doi:10.1890/12-1595.1)
60. Minor ES, Gardner RH. 2011 Landscape connectivity and seed dispersal characteristics inform the best management strategy for exotic plants. *Ecol. Appl.* **21**, 739–749. (doi:10.1890/10-0321.1)
61. Godfrey SS, Bull CM, James R, Murray K. 2009 Network structure and parasite transmission in a group living lizard, the gidgee skink, *Egernia stokesii*. *Behav. Ecol. Sociobiol.* **63**, 1045–1056. (doi:10.1007/s00265-009-0730-9)