Genetic diversity decreases toward species range edges

Chloé Schmidt[1,2,*], Jussi Mäkinen[3], Jean-Philippe Lessard[4], Colin J Garroway[5]

[1] Department of Biology, Dalhousie University; Halifax, B3H 4R2, Canada

[2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig; Leipzig, 04103, Germany

[3] Research Center for Ecological Change, Faculty of Biological and Environmental Sciences, University of Helsinki; Helsinki, Finland

[4] Department of Biology, Concordia University; Montréal, H4B 1R6, Canada

[5] Department of Biological Sciences, University of Manitoba; Winnipeg, R2T 2N2, Canada

*Correspondence: chloe.schmidt@dal.ca

Abstract

Changing species distributions amidst global change underscores a pressing need to better understand the processes that cause range limits. However, despite over a century of work, there is no consensus regarding whether there are generalizable contributors to the emergence of species range limits. Species' geographic distributions frequently taper without obvious physical or environmental barriers, raising a fundamental question about whether evolutionary processes contribute to the emergence of species range limits species. Theoretical models suggest that spatial variation in the strength of genetic drift can reduce the efficiency of natural selection to the point where range limits form. The extent to which this theory holds in real-world populations is uncertain. With data comprising 37,397 genotypes sampled across 59 mammal species' ranges from 1,271 sample sites, we show that effective population size and genetic diversity decline toward range edges, while genetic differentiation increases, indicating stronger drift and reduced gene flow in peripheral populations. These results demonstrate that limits to range expansion arise predictably as populations become small and isolated, which reduces the efficiency of natural selection. Our findings point to genetic drift as a general mechanism constraining range expansion, helping to explain the widespread structure of species' distributions and help explain commonly observed idiosyncratic patterns of range configuration and the organization of biodiversity.

Introduction

No species occurs everywhere. While species' range limits sometimes align with physical barriers or abrupt changes in environmental conditions, they also commonly taper off gradually. The emergence of tapering range edges is puzzling from an evolutionary perspective. Given sufficient time and genetic variation, theory predicts that populations should be capable of adapting in response to changing environments, including adaptive expansions into neighboring habitats[1–5]. Indeed, most populations experience strong selection throughout their ranges where they persist and often become locally adapted[6]. Populations can typically withstand or respond to selection pressures within their ranges. However, the ubiquity of range edges without clear physical barriers suggests that general evolutionary mechanisms limit adaptation at these margins.

Theoretical models of range edge evolution are well-developed and point to the potential importance of genetic drift as an evolutionary contributor to range edge formation[5,7–9]. This theoretical work is compelling because of its simplicity and potential generality. All populations experience genetic drift—the collective contribution of random influences on allele frequency changes from one generation to the next—which erodes genetic diversity over time in proportion to their effective population size[10]. The effective population size of a population decreases as abundance and gene flow between populations decrease, thereby increasing the strength of genetic drift. Foundational population genetic theory demonstrates that as allele frequency change becomes increasingly stochastic, the efficiency of the deterministic process of natural selection declines[11,12]. The inverse relationship between drift and the efficiency of selection is also well supported empirically in vertebrates, where, for example, signatures of selection at

protein coding genes across the genome become increasingly apparent as effective population sizes become larger[13].

In theory, all that is required for range edge formation is that effective population sizes decline toward range edges to the point where genetic drift overwhelms the efficiency of natural selection and populations become less likely to persist, adapt, and expand (Fig. 1). Genetic drift can be a general process influencing range limits because drift-mediated limits to the efficiency of natural selection are independent of the strength of selection acting on particular phenotypes. The fact that genetic drift reduces the efficiency of natural selection acting across all traits simultaneously means that drift-mediated reductions in the efficiency of selection can contribute to the formation of range edges regardless of specific traits or environmental contexts, which will vary in detail across species, populations, and through time[5,11]. It also means that we can test the contribution of drift-mediated limits to natural selection to range edge formation using simple measures of genetic drift (i.e., effective population size) estimated at multiple points across a species range[14]. Empirically, even if genetic drift is a general contributor to range edge formation, we can be certain that genetic drift will be one among many other, likely more idiosyncratic and context-specific, contributors to the emergence of range edges.

Although theoretical models strongly suggest that genetic drift is likely important for the emergence and maintenance of range edges across species, empirical evidence supporting the prevalence of this pattern in nature is limited (see [15]). We tested whether edge populations experience stronger genetic drift than populations near the range interior by quantifying effective population size, genetic diversity, and genetic differentiation at multiple locations across the ranges of multiple species. Effective population size estimates the strength of genetic drift, which

should be reflected in reduced genetic diversity. Due to increasing geographic isolation, small edge populations receiving lower gene flow should also become increasingly genetically differentiated and experience stronger drift (Fig. 1). Leveraging open-source genetic data for terrestrial mammals worldwide, we quantified these genetic metrics for 1,271 local populations comprising 37,397 individuals sampled across 59 species of mammals scattered across regions of the world (Fig. 2, S1-2, Table S1).

Results

Consistent with predictions for drift-mediated limits to selection, spatially and phylogenetically explicit Bayesian generalized linear mixed models indicated that effective population sizes declined nearer to range edges for most species, pointing to a general pattern of stronger drift at range margins (Table 1, Table S2, Fig. 3, Figs. S2-5). We also found lower genetic diversity closer to the range edge, in terms of both the abundance and evenness of alleles (allelic richness and gene diversity, respectively) in local populations (Table 1, Table S2, Fig. 3, Figs. S2-5). This is the predicted consequence of increased genetic drift. Finally, populations near edges also tended to be more genetically distinct compared to populations further from edges, suggesting that reduced effective population size at the edge is also in part due to reduced gene flow (Table 1, Table S2, Fig. 3, Figs. S2-5). Our findings of repeated, consistent patterns of spatial variation in the strength of genetic drift[5,9,16] and their consequences for the efficiency of natural selection emphasizes the potential importance of limits on the process of adaptation itself relative to the outcome of specific selection pressures[11,12,17]. That effective population sizes, genetic diversity, and gene flow declined toward range edges across a globally distributed sample of species provides compelling support for an important and generalizable role for genetic drift-mediated limits on genetic adaptation shaping range edges across species and localities.

Reconciling ecological and evolutionary perspectives on range edge formation

One of the best developed explanations for the emergence of range edges, the abundant center or central-marginal hypothesis, is agnostic about the reasons why species do not expand their geographical distribution through adaptive evolution at range margins. The abundant center hypothesis focuses on species ecology and posits that population sizes should peak in the center of the range where environmental conditions (including biotic and abiotic factors) best match species' needs, and decline toward edges as environments become less well suited to species' ecological needs[18]. However, it has become increasingly clear that the dynamics of populations in range interiors are complex and difficult to generalize[15,19–29]. Regions of high abundance within ranges often seem to be poorly aligned with geographic range centers, and species can have multiple abundance peaks across their range[19,20,25]. However, abundant center demographic patterns predict parallel variation in the intensity of drift, and thus decreased genetic diversity and increased population differentiation nearer range edges[5,9]. We explored support for the abundant center assumption with our data by fitting similar models using a local population's distance from the geographic range center rather than distance from a range edge. We found similar, but less clear results (Fig. S6), suggesting that distance from the geographic range center can act as a proxy for the range edge-based processes tested here.

We further explored the abundant center assumption by visualizing the positions of genetically diverse, well-connected core populations. From a population genetics perspective, population cores—wherever they occur in a range—should be characterized by high levels of gene flow and genetic diversity (Fig. 1). To do this, we mapped spatial genetic structure using pairwise F_{ST}[30] and population-specific F_{ST}[31,32] for the subset of species with the best geographic sampling across ranges (Fig. 4, S7, See SI Methods). Our maps of spatial genetic structure suggest that genetic core populations often encompass a large proportion of a species range, suggesting that

range interior dynamics tend not to be well-characterized by a single, abundant, centrally located core population.

The abundant center concept also contributes to an alternative evolutionary explanation for range edge emergence known as genetic swamping. Due to their higher abundance, central populations could be net producers of immigrants, and edge populations net recipients, generating density-dependent dispersal toward range edges[33]. Genetic swamping refers to the idea that migrants to edge populations that are locally adapted to central environmental conditions would reduce the mean fitness in edge populations and limit adaptive expansion into neighboring environments. A growing consensus supported by theory, experimentation, and observational data suggests that central-marginal patterns of gene flow tend to have neutral[22] or net positive effects on the mean fitness of edge populations due to added genetic diversity[22,34]. This suggests that rather than contributing to the emergence of range edges, gene flow to range margins typically facilitates adaptive evolution by decreasing the strength of drift. Our finding of increased genetic differentiation at range edges, which implies a restricted flow of individuals and alleles from core populations, suggests that gene flow from cores to margins is generally limited, providing additional evidence suggesting genetic swamping does not generally constrain species ranges.

Weak support for the assumption that cores occur in the geographical center of ranges, and our finding that genetic core populations are geographically large, has likely contributed significantly to the mixed support for abundant center patterns in the literature, particularly for abundance. Tests of the abundant center hypothesis are data intensive, and syntheses of this topic have had to contend with methodological heterogeneity across studies that add statistical noise to formal meta-analyses[15,20,25,26,35,36]. We were able to overcome these drawbacks by harvesting publicly available genetic data, enabling us to consistently estimate metrics of genetic composition and

range position across disparate studies. A prominent reason why the central-marginal hypothesis has been so debated stems from differences in how range position is defined. Several continuous and categorical metrics have been used to describe range position, including distance from the range edge, the geographic range center, or the environmental niche center[15,20,24]. For genetic diversity in particular, studies using continuous measures of peripherality are rare[15]. However, our findings that core populations tend to be large suggest that it is difficult to consistently categorize range position across species. Distance from the range edge, as used here, is a continuous metric that makes no assumption about where in the range the core population lies.

We focused on mammals, and our dataset had a high concentration of samples from North America and Europe, which may affect the generalizability of our results. The predictions of the theoretical model of range edge formation we test here do not vary depending on location or species characteristics—indeed, its wide applicability is necessary for understanding whether genetic drift could be a general contributor to range edge formation. Very few species deviated from the pattern of decreased genetic diversity nearer range edges in our results (Fig. 3), and there was no tendency across genetic metrics for these species to be those sampled outside of North America or Europe. Previous work based on literature surveys, categorical measures of geographic peripherality, and distance from range centroid report find generally high support for reduced genetic diversity in range edge populations, however with variable consistency of the pattern across plants and animals[15,24,37,38]. The evolutionary histories and contemporary demography of individual species can contribute to variation in spatial genetic patterns in idiosyncratic ways that would introduce noise, but not bias into statistical models[27,37]. Thus, while our data precludes full generalization across taxa and location, there is no strong reason to

suspect consistent deviation from this pattern based on species traits or characteristics of species ranges.

Genetic drift and range dynamics in a changing world

Our empirical results align with theoretical models suggesting that the interaction between environmental gradients and the evolutionary process of genetic drift combine to produce expansion thresholds at the population level that can form distributional limits for species[5,14]. However, the current most widely used tool to understand and predict range dynamics is species distribution modeling[39], which associates species occurrences with environments to infer suitable habitats for species as a whole. Using future climate scenarios, species distribution models are used to make predictions about how ranges will shift, expand, or contract in response to changes in the availability of suitable habitat. Species distribution models thus typically predict range dynamics as a direct outcome of species-level habitat associations while not explicitly considering evolutionary processes (see [40,41] for static species distribution models, but [42,43] for non-stationary species distribution models). Microevolutionary processes, which act on populations and can vary across species' ranges, have been shown to be important for predicting changes in species distributions[44–48]. Because expansion thresholds apply to populations rather than species, range dynamics in response to climate change should be highly variable and driven by intraspecific variation in population structure and environmental gradients across species ranges. For example, rapid reduction of population size and population fragmentation can cause populations to drop below the size needed for ongoing adaptation at any position in a range—not only near edges—with range-wide effects[5]. Conversely, deteriorating environmental conditions in core populations may increase dispersal to range edges, making genetic diversity available in edge populations and placing them above expansion thresholds[5]. Although poleward range shifts

198 were typically predicted to occur as species track their environmental niche, empirically, there is
199 substantial variation in range change dynamics. Poleward shifts account for less than half of
200 documented range movements[49,50], and asymmetric changes in abundance recorded across
201 species ranges during range shifts oppose predictions for non-evolutionary, species-level
202 environmental tracking, and suggest population-level processes underlie range change
203 dynamics[51].

204 Establishing how microevolutionary processes contribute to macro-scale patterns in biodiversity
205 has remained a longstanding challenge (reviewed in [52,53]). Species ranges are the unit of
206 organization for higher-level biodiversity patterns, such as biodiversity hotspots, biogeographic
207 regions, and the latitudinal richness gradient. If drift-mediated limits to range expansion stand
208 scrutiny, they could be fundamental to our ability to forecast changes in macroecological patterns
209 of biodiversity. Our results highlight a general mechanism by which contemporary, micro-scale,
210 population-level demographic and evolutionary processes can contribute to the evolution of
211 species range limits and their dynamics under global change.

212

213 **Methods**

214 Data

215 *Genetic data.* We systematically harvested publicly archived genetic data for mammal species
216 sampled across the globe to test our questions about the contributions of population genetics to
217 range edge formation. We expanded a database of raw microsatellite genotypes for mammal
218 species from Canada and the USA, initially compiled between 2017-2021[54,55], to cover archived
219 data for mammals globally. For detailed methods of dataset assembly methods see Schmidt et al.
220 [54]. Briefly, we obtained a list of 5216 names of species with ranges occurring outside of the

United States and Canada from the IUCN database to use as keywords in addition to a 'microsat*' search term (e.g., *Alces alces microsat**). We then performed a systematic search of the Dryad Digital Repository (https://datadryad.org/stash) by accessing the Dryad API using a custom R [56] script. This produced 269 results which we then manually checked for suitability, ensuring genotype data were available for the correct species, samples were collected from natural wildlife populations, the study used neutral microsatellite markers, sample design was appropriate for estimating normal levels of genetic diversity and differentiation (e.g., samples were not taken after a disturbance or recent bottleneck, if parentage analysis, offspring were excluded), and with sampling locality and population grouping information (Data S1).

For each sample location (local population), we recorded latitude and longitude in decimal degrees, year of sampling if applicable, and sequence of sampling (for populations sampled over time at the same location). We maintained local population groupings as delineated in the original work, and make no assumption about genetic populations or population structure. We recorded the number of loci and individuals sampled from the data directly in R. We retained local populations with at least 5 individuals sampled, and species with at least 5 local populations. The median number of individuals sampled per site across all species was 22 (range: 5 – 1050 individuals; Table S1). We retained only the most recent time point for the few species with temporal sampling.

We estimated four metrics of genetic composition for each local population. First, we estimated gene diversity, a measure of evenness which gives the probability of randomly sampling two different alleles from a group of individuals ('Hs' function in the *adegenet* package[57]. Because gene diversity is computed from allele frequencies and not counts of alleles, it is not strongly affected by sample size[58]. Allelic richness was our second measure of genetic diversity. We

estimated rarefied allelic richness ('allelic.richness' in *hierfstat*[59] using 10 alleles as the minimum sample size to ensure estimates were comparable across the entire dataset. We used population-specific F_{ST} as a measure of genetic differentiation ('betas' in *hierfstat*). This metric is described in more detail in the *Mapping population structure* section below. Finally, we estimated the contemporary effective population size of the parental generation for each local population using the linkage disequilibrium method implemented in the NeEstimator software (version 2.1[60] accessed from R with the rLDNe package (https://github.com/zakrobinson/RLDNe). The linkage disequilibrium method implemented in NeEstimator is among the more accurate for estimating contemporary effective population size [61], and it performs especially well for small populations. However, an estimate of infinity is returned when effective population sizes are extremely large, or when small sample size makes signals of genetic drift difficult to detect. We treated these estimates as NA values and excluded them from analysis.

Range data. We downloaded range maps from the IUCN (https://www.iucnredlist.org/resources/spatial-data-download). We filtered species range data based on presence, origin, and seasonality, retaining portions of the range where species populations were identified as extant, native, and resident. Many species ranges were fragmented, represented by multiple disjunct polygons. Measuring the distance of a genetic sample site to the nearest edge of a polygon thus did not necessarily represent a site's position with respect to the center and margin across an entire range. Instead, we measured the distance from each site to: 1) the nearest edge of a convex hull encompassing the species range, and 2) the geographic range centroid. We drew convex hulls and calculated centroids in the *terra* package, ensuring centroids were located within range polygons. We note that this restriction on range centroids to be inside polygon boundaries means they are not necessarily true centroids. Two

268 species, brown bears *Ursus arctos* and moose *Alces alces,* had ranges across North America and
269 Eurasia. In these cases, we treated range polygons across continents separately and drew convex
270 hulls and centroids for the range on each continent. We measured the geodesic distance, the
271 shortest distance between two points along a curved surface, using the *geosphere* package in R
272 [62].

273 Although we excluded introduced populations in our genetic data where identifiable and
274 removed portions of species ranges classified by the IUCN as *introduced*, *re-introduced*,
275 *vagrant*, *uncertain*, or *assisted colonization,* four species had sites (22 sites in total) located
276 outside the species' convex hull. We chose to retain these sites, as they may represent expansion
277 not yet recorded in IUCN range data and we predicted they should nevertheless adhere to
278 interior-marginal patterns. We made these distances negative to maintain the directionality and
279 interpretation of our distance measure (higher values of distance indicate moving toward the
280 range interior).

281

282 Bayesian hierarchical models

283 We modeled the genetic diversity and genetic differentiation in local populations as a function of
284 range position using a series of four Bayesian generalized linear mixed models (GLMMs; 1 per
285 genetic response variable). All response and explanatory variables were scaled and centered so
286 that models were comparable. We expected substantial variation in mean values of genetic
287 composition variables across species, and variation in the effect of distance to range edge across
288 species[15]. We captured this structure in our models using random effects that allowed intercepts
289 and slopes to vary across species. For each response variable we defined the linear predictor as:

290 $$y_j(s_i) = \alpha + x(s_i)\beta + \eta_j + x(s_i)\gamma_j + h_j + g(s_i) + \varepsilon_i$$

where $y_j(s_i)$ is response value of species j at location s of sample i, α is the model intercept, $x(s_i)$ is distance to range edge, β is a fixed effect of range edge, η_j is the species-specific intercept, γ_j is species specific deviation from the fixed effect of range edge, h_j is a phylogenetically correlated species specific intercept, $g(s_i)$ is a spatial random effect, and ε_i are the model residuals. With this model structure, we could estimate the average response (β) of a genetic response variable to the distance to range edge and the random variation between species in how their responses (γ_j) deviate from the average. To account for the conspecific variation in the average of a genetic metric we included independent (η_j) and phylogenetically (h_j) correlated species-level random intercepts. Lastly, the locations of sample sites may cause model residuals to be spatially autocorrelated, which violates the assumptions of linear regression. We thus accounted for spatially varying intercepts by testing for spatial autocorrelation and including a spatial random effect $g(s_i)$ when necessary. Our motivation for using a spatial random effect was only to explain the spatially correlated residual variation, not to further explore spatial variation.

We defined the random effects as follows:

The species-level random intercept and response to range edge were defined as two-dimensional correlated random effects, which are independently distributed among species. The effects are defined as:

$$[\eta, \gamma] \sim N(0, W^{-1}),$$

where ${W^{-1}}_{\eta,\eta} = 1/\tau_\eta$, ${W^{-1}}_{\gamma,\gamma} = 1/\tau_\gamma$ and ${W^{-1}}_{\eta,\gamma} = \rho/\sqrt{(\tau_\eta \tau_\gamma)}$. The precision parameter τ captures the variance of the species-specific effects, and ρ captures their correlation. Hence, we test whether species that, on average, have higher values of a genetic response variable have a

larger or smaller effect of distance to range edge than an average species. In the case that ρ deviates from zero, we can borrow strength across the random effects and narrow their conditional uncertainties. This method has considerable advantages over modeling each species separately because it reduces Type I error rates, and partial pooling in hierarchical models allows less well-sampled species to borrow statistical strength from better sampled species, resulting in better estimates of effect sizes. We assigned a Wishart-distribution prior with four degrees of freedom, marginal variances set to one, and correlation to zero.

We defined the phylogenetic random effect as a generic random effect so that

$$[h] \sim N(0, {Q_h}^{-1}),$$

where ${Q_h}^{-1} = \tau_h C$. The inverse correlation matrix C is the inverse of the phylogenetic distance matrix computed with the function inverseA from the R-package *MCMCglmm*[63]. The variance parameter τ_h was assigned a penalized complexity prior which gives a 0.1 prior probability that the standard deviation of h (= ${\tau_h}^{-1/2}$) has a value larger than one. This is a relatively vague prior. We also fit models without a phylogenetic random effect to test the sensitivity of our results to the inclusion of the phylogeny (Table S3).

Lastly, we defined the spatial random effect as a Besag-type random effect, for which the value in a location is an average of the values of the adjacent sampling locations[64]. The random effect at location *s* conditional on connected study location is defined as

$$g(s_i)|g(s_{-i}) \sim N(\mu(s_i), {Q_g}^{-1}),$$

where $\mu(s_i)$ is the average over the g in the nearby study locations and ${Q_g}^{-1} = 1/N\tau_g$. N is the number of connected study locations and τ_g is the precision parameter. We assigned it a penalized complexity prior which gives a 0.1 prior probability that the standard deviation of g (=

$\tau_g^{-1/2}$) has a value larger than one. We defined the connectivity structure of the study sites with a k-nearest neighbor method by fixing the number of neighbors per site to 4 for effective population size, 6 for population-specific F_{ST}, 8 for gene diversity, and 2 for allelic richness. We chose these connectivity structures by running model candidates with numbers of connections per site varying from 1 to 8 and comparing them to models with no spatial random effect. The k-nearest neighbor is an unsupervised algorithm which assigns connections between sites according to *a priori* rules. It considers only the geographical distances between sites and does not consider oceans or biogeographic regions which could affect the genetic metrics. We visually inspected all candidate connectivity structures to ensure they captured the neighborhood structures on each continent and avoided between-continent connections. We measured spatial autocorrelation with the probability that the Moran Index of residuals deviated from zero, and excluded models with spatially correlated residuals. Of the remaining model candidates, we chose the model with the best information content according to WAIC[65] (Table S4).

We ran parallel analyses with the same model structure and selection procedure with distance to range centroid as an explanatory variable instead of distance to range edge. We fit all models in R version 4.1.2 using the integrated nested Laplace approximation with the INLA package[66]. We computed marginal and conditional R^2 values for mixed-effects models as additional indicators of model fit (Table S2)[67,68]. Marginal and conditional R^2 partition the variance explained by a model into the component explained by fixed effects only (marginal R^2) and that explained by both fixed and random effects (conditional R^2).

Mapping population structure

We explored spatial patterns of genetic differentiation for species using an approach from Kitada et al.[32] that combines estimates of population-specific F_{ST} and pairwise F_{ST} to characterize spatial population structure. Population-specific F_{ST}[31] is a measure of population divergence estimated for individual samples—not pairs of samples—that uses proportions of matching alleles for pairs of alleles drawn from within and between populations to estimate differentiation. Population-specific F_{ST} describes the decay in genetic diversity and similarity from the most diverse local population, which is what we were interested in mapping and modeling. Because it is an estimate of differentiation relative to a single, most diverse sample, population-specific F_{ST} does not provide information about how differentiated sample sites are relative to each other. For this, we use pairwise F_{ST}. Combining population-specific F_{ST}, pairwise F_{ST}, and spatial metadata can provide a general idea about regions with large populations well-connected by gene flow that may represent the "core" areas of a species range. We estimated both F_{ST} metrics in the hierfstat package, using the 'pairwise.neifst' function to estimate Nei and Chesser's pairwise G_{ST}[69], and 'betas' to estimate population-specific F_{ST}.

To visualize patterns of genetic diversity decay and genetic differentiation, we generated 2 plots per species to represent population structure in complementary ways (Fig. 4, Fig. S7); for a detailed description of these methods see [32]. We focused on 12 species that had good coverage of their range. The first plot showed population-specific F_{ST} values across species' ranges. To visualize the major axes of genetic differentiation, we performed a principal coordinates analysis (PCoA) on the pairwise genetic differentiation matrix and plotted site positions on the first two PCoA axes.

Acknowledgments: We thank Sean Robertson, Leah Kathan, Jessie Ogden, and Meghan Mahoney for early discussions on this topic. We also thank the Population Ecology and Evolutionary Genetics group and the Evolutionary Biology BIOL 3300 students at the University of Manitoba for their comments and feedback on manuscript drafts. C.S. acknowledges the support of iDiv funded by the German Research Foundation (DFG–FZT 118, 202548816) and a Natural Sciences and Engineering Research Council of Canada Discovery Grant. C.G. was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant.

Author contributions:

Conceptualization: CS, CJG, JPL

Methodology: CS, JM, CJG

Investigation: CS

Formal analysis: CS, JM

Visualization: CS

Writing – original draft: CS, CG

Writing – review and editing: all authors

Funding acquisition: CS, CJG

Data and code availability: All data underlying this work are publicly available (Table S1). Analysis code is available at: https://github.com/chloewsch/range_edge_genetics.

References

1. Antonovics, J. The Input from Population Genetics: 'The New Ecological Genetics'. *Systematic Botany* **1**, 233–245 (1976).
2. Bradshaw, A. D. Genostasis and the Limits to Evolution. *Philosophical Transactions: Biological Sciences* **333**, 289–305 (1991).
3. Bridle, J. R. & Vines, T. H. Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology and Evolution* **22**, 140–147 (2007).
4. Gaston, K. J. *The Structure and Dynamics of Geographic Ranges*. (Oxford University Press, Oxford, 2003).

5. Polechová, J. Is the sky the limit? On the expansion threshold of a species' range. *PLOS Biology* **16**, e2005372 (2018).
6. Gillespie, J. H. *The Causes of Molecular Evolution*. (Oxford University Press, New York, 2002).
7. Polechová, J., Barton, N. & Marion, G. Species' range: adaptation in space and time. *The American Naturalist* **174**, E186-204 (2009).
8. Kirkpatrick, M. & Barton, N. Evolution of a species' range. *The American Naturalist* **150**, 1–23 (1997).
9. Polechová, J. & Barton, N. H. Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 6401–6406 (2015).
10. Waples, R. S. What is Ne, anyway? *Journal of Heredity* **113**, 371–379 (2022).
11. Ohta, T. The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics* **23**, 263–286 (1992).
12. Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press, Cambridge, 1983).
13. Marino, A., Debaecker, G., Fiston-Lavier, A.-S., Haudry, A. & Nabholz, B. Effective population size does not explain long-term variation in genome size and transposable element content in animals. *eLife* **13**, RP100574 (2025).
14. Connallon, T. & Sgrò, C. M. In search of a general theory of species' range evolution. *PLoS Biology* **16**, 1–6 (2018).
15. Eckert, C. G., Samis, K. E. & Lougheed, S. C. Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Molecular Ecology* **17**, 1170–1188 (2008).

16. Bridle, J. R., Kawata, M. & Butlin, R. K. Local adaptation stops where ecological gradients steepen or are interrupted. *Evolutionary Applications* 1–14 (2019) doi:10.1111/eva.12789.
17. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195–205 (2009).
18. Brown, J. H. On the Relationship between Abundance and Distribution of Species. *The American Naturalist* **124**, 255–279 (1984).
19. Dallas, T., Decker, R. R. & Hastings, A. Species are not most abundant in the centre of their geographic range or climatic niche. *Ecology Letters* **20**, 1526–1533 (2017).
20. Fristoe, T. S., Vilela, B., Brown, J. H. & Botero, C. A. Abundant-core thinking clarifies exceptions to the abundant-center distribution pattern. *Ecography* 1–15 (2022) doi:10.1111/ecog.06365.
21. Hargreaves, A. L., Samis, K. E. & Eckert, C. G. Are species' range limits simply niche limits writ large? A review of transplant experiments beyond the range. *American Naturalist* **183**, 157–173 (2014).
22. Kottler, E. J., Dickman, E. E., Sexton, J. P., Emery, N. C. & Franks, S. J. Draining the Swamping Hypothesis: Little Evidence that Gene Flow Reduces Fitness at Range Edges. *Trends in Ecology and Evolution* **36**, 533–544 (2021).
23. Lee-Yaw, J. A. *et al.* A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. *Ecology Letters* **19**, 710–722 (2016).
24. Pironon, S. *et al.* Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. *Biological Reviews* **92**, 1877–1909 (2017).
25. Sagarin, R. D. & Gaines, S. D. The 'abundant centre' distribution: To what extent is it a biogeographical rule? *Ecology Letters* **5**, 137–147 (2002).

26. Santini, L., Pironon, S., Maiorano, L. & Thuiller, W. Addressing common pitfalls does not provide more support to geographical and ecological abundant-centre hypotheses. *Ecography* **42**, 696–705 (2019).
27. Singhal, S., Wrath, J. & Rabosky, D. L. Genetic variability and the ecology of geographic range: A test of the central-marginal hypothesis in Australian scincid lizards. *Molecular Ecology* **31**, 4242–4253 (2022).
28. Rolland, J., Lavergne, S. & Manel, S. Combining niche modelling and landscape genetics to study local adaptation: A novel approach illustrated using alpine plants. *Perspectives in Plant Ecology, Evolution and Systematics* **17**, 491–499 (2015).
29. Ledoux, J.-B. *et al.* Potential for adaptive evolution at species range margins: contrasting interactions between red coral populations and their environment in a changing ocean. *Ecology and Evolution* **5**, 1178–1192 (2015).
30. Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press, New York Chichester, West Sussex, 1987).
31. Weir, B. S. & Goudet, J. A Unified Characterization of Population Structure and Relatedness. *Genetics* **206**, 2085–2103 (2017).
32. Kitada, S., Nakamichi, R. & Kishino, H. Understanding population structure in an evolutionary context: Population-specific FST and pairwise FST. *G3: Genes, Genomes, Genetics* **11**, (2021).
33. Haldane, J. B. S. The relation between density regulation and natural selection. *Proceedings of the Royal Society of London. Series B - Biological Sciences* **145**, 306–308 (1956).
34. Polechová, J. The costs and benefits of dispersal in small populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**, 2021.12.16.472951 (2022).

35. Sagarin, R. D., Gaines, S. D. & Gaylord, B. Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends in Ecology and Evolution* **21**, 524–530 (2006).
36. Soberón, J., Peterson, A. T. & Osorio-Olvera, L. A comment on "Species are not most abundant in the centre of their geographic range or climatic niche". *Rethinking Ecology* **3**, 13–18 (2018).
37. De Kort, H. *et al.* Life history, climate and biogeography interactively affect worldwide genetic diversity of plant and animal populations. *Nature Communications* **12**, 516 (2021).
38. Lira-Noriega, A. & Manthey, J. D. Relationship of genetic diversity and niche centrality: A survey and analysis. *Evolution* **68**, 1082–1093 (2014).
39. Guisan, A. & Thuiller, W. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**, 993–1009 (2005).
40. A. Lee-Yaw, J., L. McCune, J., Pironon, S. & N. Sheth, S. Species distribution models rarely predict the biology of real populations. *Ecography* **2022**, 1–16 (2022).
41. Zurell, D., Jeltsch, F., Dormann, C. F. & Schröder, B. Static species distribution models in dynamically changing systems: how good can predictions really be? *Ecography* **32**, 733–744 (2009).
42. Benito Garzón, M., Alía, R., Robson, T. M. & Zavala, M. A. Intra-specific variability and plasticity influence potential tree species distributions under climate change. *Global Ecology and Biogeography* **20**, 766–778 (2011).
43. Hällfors, M. H. *et al.* Addressing potential local adaptation in species distribution models: implications for conservation under climate change. *Ecological Applications* **26**, 1154–1169 (2016).

44. Bay, R. A. *et al.* Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83–86 (2018).
45. Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y. & Keller, S. R. Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources* **21**, 2749–2765 (2021).
46. Jay, F. *et al.* Forecasting changes in population genetic structure of alpine plants in response to global warming. *Molecular Ecology* **21**, 2354–2368 (2012).
47. D'Amen, M. & Azzurro, E. Integrating univariate niche dynamics in species distribution models: A step forward for marine research on biological invasions. *Journal of Biogeography* **47**, 686–697 (2020).
48. Hanspach, J., Kühn, I., Schweiger, O., Pompe, S. & Klotz, S. Geographical patterns in prediction errors of species distribution models: Patterns in prediction error. *Global Ecology and Biogeography* **20**, 779–788 (2011).
49. Platts, P. J. *et al.* Habitat availability explains variation in climate-driven range shifts across multiple taxonomic groups. *Scientific Reports* **9**, 1–10 (2019).
50. Rubenstein, M. A. *et al.* Climate change and the global redistribution of biodiversity: substantial variation in empirical support for expected range shifts. *Environmental Evidence* 1–21 (2023) doi:10.1186/s13750-023-00296-0.
51. Chaikin, S., Riva, F., Marshall, K. E., Lessard, J.-P. & Belmaker, J. Marine fishes experiencing high-velocity range shifts may not be climate change winners. *Nat Ecol Evol* **8**, 936–946 (2024).
52. Charlesworth, B., Lande, R. & Slatkin, M. A Neo-Darwinian Commentary on Macroevolution. *Evolution* **36**, 474 (1982).

53. Hancock, Z. B., Lehmberg, E. S. & Bradburd, G. S. Neo-darwinism still haunts evolutionary theory: A modern perspective on Charlesworth, Lande, and Slatkin (1982). *Evolution* **75**, 1244–1255 (2021).
54. Schmidt, C., Domaratzki, M., Kinnunen, R. P., Bowman, J. & Garroway, C. J. Continent-wide effects of urbanization on bird and mammal genetic diversity. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20192497 (2020).
55. Schmidt, C. & Garroway, C. J. Systemic racism alters wildlife genetic diversity. *Proceedings of the National Academy of Sciences of the United States of America* **119**, 0–3 (2022).
56. R Core Team. R: A Language and Environment for Statistical Computing. (2021).
57. Jombart, T. *et al.* adegenet: Exploratory Analysis of Genetic and Genomic Data. (2017).
58. Charlesworth, B. & Charlesworth, D. *Elements of Evolutionary Genetics*. (Roberts & Company Publishers, Greenwood Village, Colorado, USA, 2010).
59. Goudet, J. & Jombart, T. hierfstat: Estimation and Tests of Hierarchical F-Statistics. (2015).
60. Do, C. *et al.* NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **14**, 209–214 (2014).
61. Gilbert, K. J. & Whitlock, M. C. Evaluating methods for estimating local effective population size with and without migration. *Evolution* **69**, 2154–2166 (2015).
62. Hijmans, R. J. geosphere: Spherical Trigonometry. (2019).
63. Hadfield, J. MCMC GLMM. (2018).
64. Besag, J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society: Series B (Methodological)* **36**, 192–225 (1974).
65. Gelman, A., Hwang, J. & Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat Comput* **24**, 997–1016 (2014).

66. Lindgren, F. & Rue, H. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software* **63**, 1–25 (2015).
67. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133–142 (2013).
68. Johnson, P. C. D. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution* **5**, 944–946 (2014).
69. Nei, M. & Chesser, R. K. Estimation of fixation indices and gene diversities. *Annals of Human Genetics* **47**, 253–259 (1983).

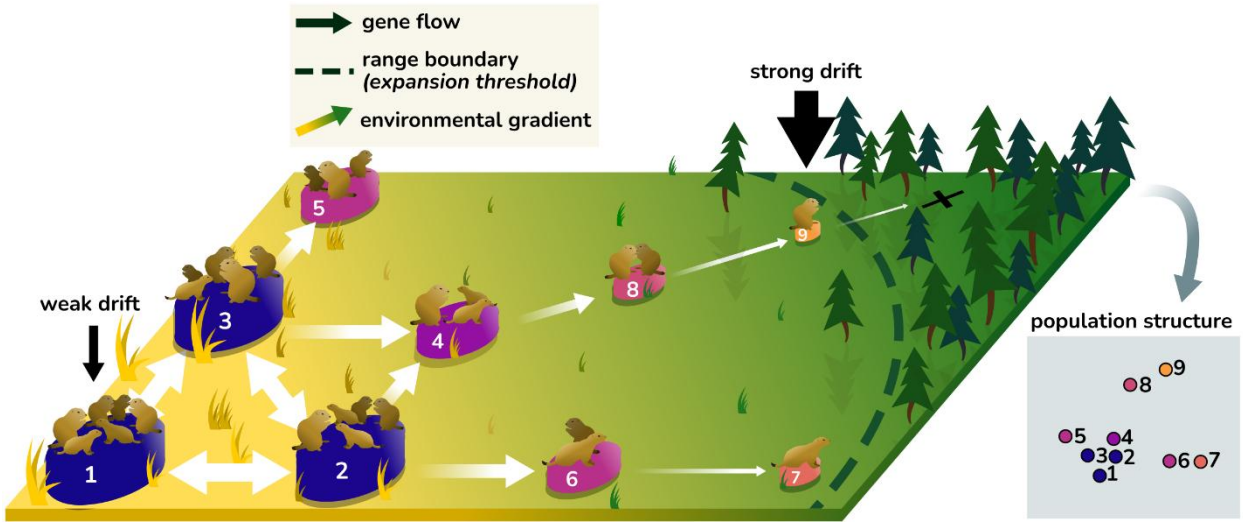


Fig. 1. Hypothesized conceptual model of drift-limited range expansion. Population sizes are larger and more well-connected toward the range interior where environments are more suitable (populations 1-3). As the environment changes, populations become smaller, more isolated, and more genetically differentiated (populations 4-9). Expansion thresholds form when populations are so small that genetic drift overwhelms natural selection and adaptation is no longer possible (populations 7 and 9).

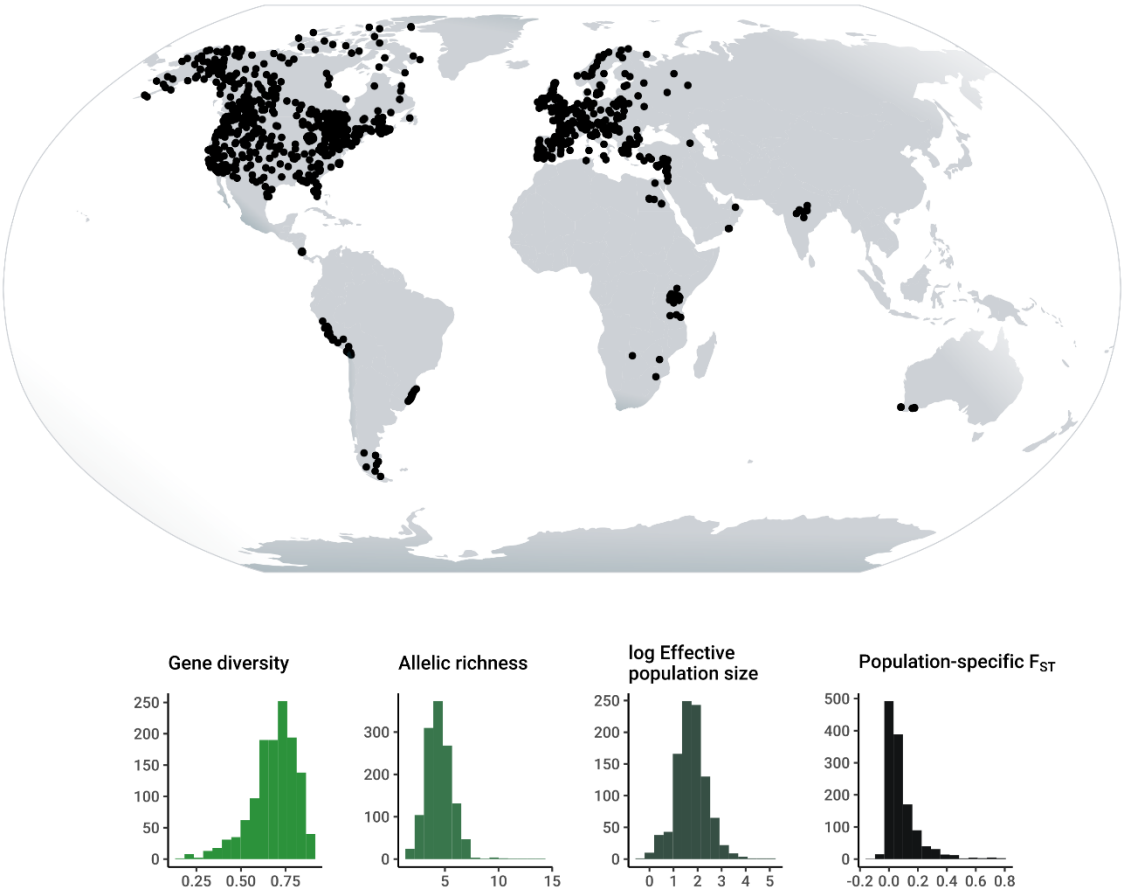


Fig. 2. Overview of genetic data. The map shows the spatial distribution of genetic data from 1271 sample locations of 59 mammal species (see species-specific maps in Fig. S1). Histograms show the distribution and range of values for each metric of genetic composition: gene diversity and allelic richness are metrics of diversity, effective population size measures the strength of genetic drift, and population-specific F_{ST} (fixation index) is a measure of genetic differentiation.

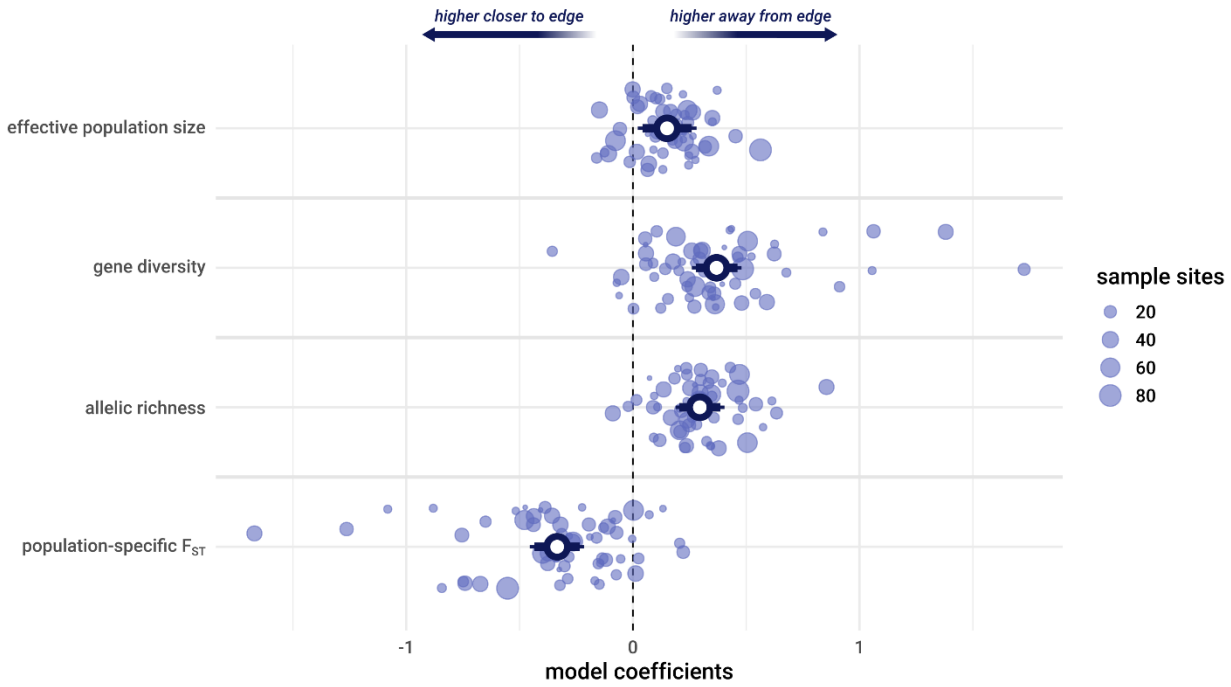
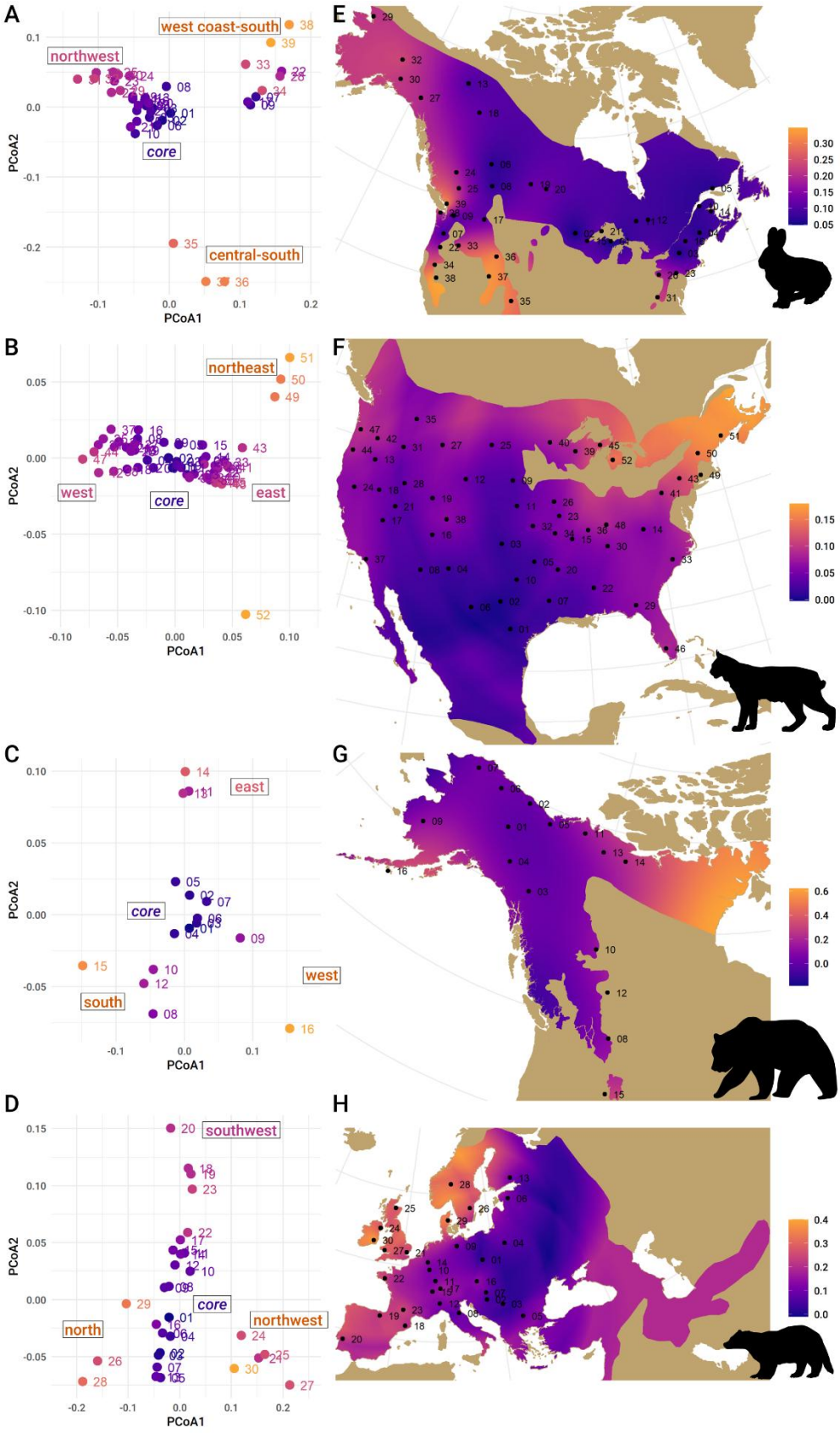


Fig. 3. Variation in genetic diversity and differentiation across species ranges. Model coefficients of the relationship between range position (distance to the edge of range convex hull) and each genetic metric (y axis) from Bayesian generalized mixed models controlling for the spatial distribution of sites and species relatedness. Effective population size is a contemporary effective size estimated using the linkage disequilibrium method for single-samples (SI Methods). Gene diversity and allelic richness (rarefied) are two metrics of genetic diversity. Population-specific F_{ST} is a measure of genetic differentiation that estimates how far single populations have diverged from a common ancestor of all sites in the sample. Open circles represent the relationship between range position and genetic metrics across all species with 90% (thin lines) and 95% (thick lines) credible intervals. Filled circles are estimated species-specific relationships between range position and the genetic metrics. The dashed vertical line indicates an effect size of 0. Positive effect sizes indicate genetic diversity and connectivity increase with increasing distance from the range edge, while negative effect sizes mean genetic diversity is reduced and local populations are more differentiated closer to the range edge.



590 **Fig. 4. Patterns of range-wide population differentiation.** Spatial population structure is

mapped for species with whole range coverage: snowshoe hare (*Lepus americanus*), bobcat (*Lynx rufus*), grizzly bear (*Ursus arctos*), and Eurasian badger (*Meles meles*); see Fig S4 for additional species. Colors represent population-specific F_{ST} values, with darker blue colors depicting local populations that are the most genetically similar to the most diverse local population compared to orange colored sites, which are the most strongly genetically differentiated. Plots (**A-D**) show major axes of population differentiation from a principal coordinates analyses (PCoA) of Nei's pairwise-F_{ST}, which measures the magnitude of genetic differentiation between pairs of local populations. The central location of genetic core populations with low differentiation in PCoA plots suggest they are well-connected by gene flow. Labels indicate spatial direction of differentiation that correspond to maps in (**E-H**). Large dark blue areas across maps suggests that larger, well-connected core populations can cover large proportions of the species range and are not limited to the geographic range center.

Table 1. Summaries of Bayesian hierarchical models relating genetic metrics to range position. The effect size of range position is given with 95% credible intervals. Range position is described by the distance to the edge of the range convex hull. Models controlled for spatial and phylogenetic correlation. $\sigma_{intercept}$ is the standard deviation of species-specific intercepts, and summarizes variation in mean genetic diversity, effective population size, or F_{ST} across species. σ_{slope} is the standard deviation of species-specific estimates of the effect of distance, and shows variation in species responses to range position. Marginal log-likelihood and DIC (Deviance Information Criterion) are two model fit metrics.

Variable	Range position	$\sigma_{intercept}$	σ_{slope}	Marginal log likelihood	WAIC
gene diversity	0.37 (0.26 – 0.48)	0.79 (0.59 – 1.03)	0.48 (0.36 – 0.64)	-1912.16	1635.06
allelic richness	0.30 (0.19 – 0.40)	0.66 (0.47 – 0.93)	0.32 (0.24 – 0.44)	-1795.71	2895.87
effective population size	0.15 (0.02 – 0.28)	0.41 (0.31 – 0.53)	0.33 (0.23 – 0.45)	-1939.89	2511.31
population-specific F_{ST}	-0.33 (-0.45 – -0.22)	0.56 (0.41 – 0.78)	0.48 (0.35 – 0.64)	-2300.26	2614.05