Title: Common causes drive negative correlations between nuclear genetic and species level biodiversity

Authors: Chloé Schmidt[1*], Stéphane Dray[2], Colin J. Garroway[1*]

Affiliations:

[1]Department of Biological Sciences, University of Manitoba, Winnipeg, Canada

[2] Univ Lyon, Université Claude Bernard Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, F-69100, Villeurbanne, France

*Correspondence to: schmid46@myumanitoba.ca, colin.garroway@umanitoba.ca

Abstract: We do not yet have a solid empirical understanding of the processes that produce biogeographic patterns of species richness. This is partly due to a lack of knowledge about corresponding spatial patterns of genome-wide diversity, which will be inextricably linked to species richness. We use estimates of gene diversity calculated from open data to show that genetic diversity and species richness share spatial structure. Species richness hotspots tend to harbor low levels of within species genetic variation. Fitting multiple response and predictor variables structured as a hypothesis network showed that a single model encompassing eco-evolutionary processes related to environmental energy availability, niche availability, and proximity to humans explains 75% of variation in the gene diversity gradient and 90% of the variation in species-level diversity. This advances our understanding of the patterns and joint causes of variation in the two most fundamental products of evolution.

Keywords: more individuals hypothesis, heterogeneity, latitudinal diversity gradient, carrying capacity, Anthropocene, macroecology, macroevolution

Introduction

Biogeographical patterns in species richness are particularly well described and studied. Identifying how eco-evolutionary processes shape species gradients is important for our basic knowledge of the products of evolution and societal wellbeing, but an empirical understanding of the processes causing these gradients has remained elusive (Etienne et al. 2019). Genetic diversity is the most fundamental level of biodiversity. But until recently, comparable multi-species, continent-wide data describing genome-wide variation was not available. These two levels of diversity are so entangled that they are nearly inseparable. By not incorporating genetic diversity into analyses of species richness we have been missing a critical piece of the complete picture.

Species richness gradients

Relationships between species richness and environments suggest how we can integrate genetic diversity into a joint analysis of both levels of variation. Though described as the latitudinal species gradient, latitude can only be a correlate of the gradient's causes. Indeed, species richness patterns sometimes deviate significantly from latitudinal trends, and these deviations suggest its underlying causes. In North America (Simpson 1964) (Fig. 1) the latitudinal trend is pronounced between the Arctic and the Canada-USA border (~50°N), before disappearing across the continental USA and reappearing near its border with Mexico (~30°N). Species richness in the continental USA varies longitudinally, peaking in the west. These deviations suggest that species richness is correlated with environmental energy availability and habitat heterogeneity (Simpson 1964; Kerr and Packer 1997), which in turn are both correlated with latitude in some areas.

This pattern suggests that environments play a prominent role in determining species richness and genetic diversity in at least two important ways. First, energy availability can impose an upper limit on both the number of individuals and the number of species a given area can support (the more-individuals hypothesis) (Storch et al. 2018). Diversity tends to increase with the number of individuals in an assemblage, both in terms of the number of species in a community (Hubbell 2001), and genetic diversity within populations (Kimura 1983). Second, more complex environments have more niches. This can limit dispersal and support a greater diversity of species, albeit at smaller population sizes, if those species come to specialize on different resources (heterogeneity hypotheses) (Kadmon and Allouche 2007). As specialized populations diverge, genetic variation will be divided among species that no longer interbreed. These smaller populations will also lose genetic diversity due to genetic drift faster than large populations.

These two processes suggest that spatial variation at species and genetic levels are partly set by environmental carrying capacities that limit the number of individuals and species a given area can support. The effects of limited carrying capacity on genetic diversity at the population level are well understood: smaller populations should be less diverse than larger populations due to drift. Recent modelling suggests that carrying capacity, relative to other processes that also produce gradients, is the strongest and most stable contributor to species richness (Etienne et al. 2019; Brodie 2019).

Shared spatial variation in species and nuclear genetic gradients

To test for shared spatial variation we quantified spatial patterns of nuclear genetic diversity for North American mammals at a continental scale: the scale at which biogeographical diversity patterns occur. To do this, we repurposed raw microsatellite genotypes posted in public data

repositories (Schmidt et al. 2020b). Microsatellite markers estimate genome-wide diversity—the quantity we are interested in—well (Mittell et al. 2015). They were also the most commonly archived marker type, which allowed us to maximize sample size. Our final data set consisted of 34,841 raw genotypes from 38 mammalian species sampled across 801 sites. We calculated gene diversity, an estimate of the evenness and spread of alleles that is not particularly sensitive to sample size (Fig. S1), for each site as our measure of genetic diversity (Charlesworth and Charlesworth 2010). We then estimated species richness at each site by counting the number of native mammal species whose ranges overlapped the site (IUCN 2019) so that we had directly comparable estimates of diversity at the genetic and species levels. We summarized spatial patterns in genetic diversity and species richness using distance-based Moran's eigenvector maps (MEMs) (Dray et al. 2006). Sixty-five percent of the variation in species richness and 24% of variation in genetic diversity was spatial (Fig. S2). Variance partitioning suggested that 85% of the total spatial variation in genetic diversity, and 32% of spatial variation in species richness was accounted for by spatial patterns shared at both levels of diversity.

We used the linear regression predicted values of spatial vectors (MEMs), calculated above, that had Moran's I measures of spatial autocorrelation > 0.25 (indicating broad spatial patterns) to produce maps of continental-scale spatial variation in both levels of diversity (Fig. 1). Consistent with previously described patterns of species richness in North America, species richness hotspots at our genetic sample sites appear to be related to topography, a measure of environmental heterogeneity, and potential evapotranspiration, a measure of energy availability (Fig. 1). There was no obvious relationship between latitude and nuclear genetic diversity. Similar to patterns of species richness, a longitudinal gradient in genetic diversity is the dominant pattern for North American mammals—however, diversity gradients at the two levels

trend in opposite richness. Genome-wide genetic diversity appears markedly lower in regions with high species richness, such as on the west and mid-Atlantic coasts, where there is high energy availability and topographic relief. This pattern is consistent with the prediction that these areas would support more species with smaller, less genetically diverse populations. Our map of neutral nuclear genetic diversity contrasts with relatively consistent latitudinal gradients in mitochondrial (mtDNA) genetic diversity, the only other genetic marker explored at a similar spatial scale (Miraldo et al. 2016; Theodoridis et al. 2020). This is perhaps not surprising because mtDNA diversity is not systematically related to population size (Bazin et al. 2006), species ecology, or life history (Nabholz et al. 2008)—all key features of the processes we explore here.

Common causes of species and nuclear genetic gradients

To explore the common causes of genetic and species-level diversity, we built a conceptual model based on the idea that carrying capacity is a strong driver of diversity. We fit this conceptual model to data using structural equation modelling, an approach for examining cause-effect relationships within hypothesis networks that can accommodate multiple predictor and response variables. Structural equation modelling is an extension of multivariate multiple regression where variables can be thought of as nodes in a network, and directional paths connecting nodes represent causal relationships. The strengths of paths are equal to regression coefficients (Shipley 2016). In addition to direct effects, you can quantify indirect effects between variables by multiplying direct effects over paths. Using standardized coefficients, we can compare the strength of relationships both within and across levels of biodiversity. The appropriateness of links in the hypothesis network can be tested using tests of directed separation (Shipley 2016), where the null hypothesis is that the two variables are independent, conditional

116 on other predictors of either variable. This means that although we start with a focus on carrying
117 capacity, the data can suggest the addition or removal of links representing alternative
118 hypotheses.
119
120 Our conceptual model was built around the predicted effects of carrying capacity related to the
121 more individuals and environmental heterogeneity hypotheses (Fig 3a). The more individuals
122 hypothesis predicts that increasing energy availability should act through population size to
123 increase both species and genetic diversity. We use body size, which is inversely related to
124 population size, as our estimate of species-level population size (Damuth 1981). We measured
125 energy availability as the mean annual potential evapotranspiration across a species range.
126 Heterogeneity, measured at the range level by quantifying the area-corrected range in elevation
127 across a species range, should reduce population sizes but lead to greater species richness by
128 increasing niche diversity. Heterogeneity and energy availability were measured at the range-
129 level because the spatial coverage of genetic sample sites in the data is not evenly distributed.
130 Additionally, some species ranges could be oversampled if we considered population-level
131 environmental variation and thus overrepresented compared to species ranges that contain fewer
132 sampled populations. Finally, we included human population density in our model, predicting
133 negative relationships with species richness, genetic diversity, and mass (Merckx et al. 2018).
134 Contemporary rapid environmental change is rarely considered at the same time as long-term
135 processes, but humans are known to influence both levels of biodiversity and so should be
136 modelled.
137
138 Our final model fit the data well (SEM p= 0.23, Fisher's C= 2.92; Fig. 3b, Table S1). Energy
139 availability, niche heterogeneity, and human population density, acting both directly and

indirectly through species population size, explained 32% of the variation in genetic diversity. The species-level variation explained by the random effect for species brought the total variation in genetic diversity explained by our model to 75%. The same model explained 90% of the variation in species richness. There was no strong spatial autocorrelation in the model, suggesting that the spatial structure of the diversity data was well captured by our model covariates (Fig. S3).

All links in our conceptual model were supported, with additional direct links suggested from energy availability to species richness, genetic diversity to species richness, and heterogeneity to genetic diversity (Fig 3b). Mammals conformed to the zero-sum carrying capacity related expectations of the more individuals hypothesis at both genetic and species levels of biodiversity. Our data indicated that when resources are limited, environmental carrying capacity is limited. These environments supported fewer but larger-bodied species with smaller population sizes and lower genetic diversity. In resource-rich areas, organisms are generally smaller, and populations and communities larger, harboring greater genetic diversity and species richness. The strength of effects related to the more individuals hypothesis was most prominent at the genetic level of diversity: the strength of the indirect effect of energy on genetic diversity acting via population size was 0.13 compared to 0.02 for species richness. We also detected a direct effect of energy on species richness (path coefficient = 0.44 ± 0.01 SE; Fig. 3b, Table S1), even after accounting for population size and topographic heterogeneity. This relationship has been noted elsewhere and has sometimes been interpreted as refuting the more individuals hypothesis (Storch et al. 2018). Vegetation structure may drive the link between species richness and temperature (Pautasso and Gaston 2005; Jiménez-Alfaro et al. 2016), as complex, vegetation-rich habitats in warmer environments also have greater niche availability. Because both links are retained in our

model it seems clear that this additional link does not negate the more individuals hypothesis, but rather is additive and indeed more important in determining species richness than the more individuals effect.

Heterogeneity was the strongest single predictor of species richness (path coefficient = 0.70 ± 0.01), and a good predictor of genetic diversity (path coefficient = -0.30 ± 0.07). Directions of effects were as expected if niche diversity reduces population sizes, carving existing variation into multiple species and leading to increased drift (Fig. 2). Because gene diversity is not a measure of divergence, we also tested whether environmental heterogeneity predicted evolutionary divergence at the population level. To do this we calculated a population-specific F_{ST} (Weir and Goudet 2017) from raw genotypes and related it to environmental heterogeneity. Population-specific F_{ST} can be interpreted as a relative estimate of the time since a population has diverged from a common ancestor. Results from this linear mixed model while controlling for species as a random effect and spatial structure, showed that heterogeneity indeed increased population divergence ($\beta = 0.13 \pm 0.06$ SE, n = 785 sites), suggesting that genetic drift is strong and gene flow limited in these areas. Heterogeneous environments impose greater spatially varying selection and coupled with low gene flow this creates ideal conditions for local adaptation, which can happen even under relatively high levels of genetic drift (Hämälä et al. 2018). This lends support to the idea that there are higher diversification rates in more complex environments because there are more opportunities for speciation.

Recent human-caused environmental disturbance both directly and indirectly (via body mass/population size) affected both species and genetic diversity (Fig. 2b). Notably, although it seems human presence and heterogeneity both reduce genetic diversity and limit dispersal,

188 human-dominated environments do not yet appear to be creating opportunities for coexistence by
189 niche-packing. Perhaps this will change with time, but if conditions for energetic requirements
190 and niche variability are not met, such a scenario is unlikely. Although a subset of species do
191 well in cities, the broader effects of habitat loss and homogenization make cities, as they are
192 currently built, inhospitable substitutes for the variety of natural niches they replace.
193
194 **Discussion**

195 The latitudinal species richness gradient has been recognized since the 1800s (Willig et al. 2003).
196 The gradient's consistency has generated >30 hypotheses aiming at explaining its relationship
197 with environments (Pontarp et al. 2019). Hypotheses fall into three broad categories:
198 evolutionary time, diversification rates, and ecological limits (differential carrying capacities),
199 which are often, at least implicitly, treated as competing ideas. But we cannot have speciation
200 without ecology and evolution. Pontarp and Wiens (2017) advocate a more interconnected view,
201 reporting that time for speciation effects on diversity should be strongest at short time scales,
202 especially during the initial colonization of new environments. When all locales are colonized,
203 habitats that provide more opportunities for speciation should over time become the most
204 diverse. All the while, differential carrying capacities determine demographic parameters such as
205 colonization success, population viability, and risk of local extinction, as well as the efficiency of
206 selection. Etienne et al. (2019) used simulations to determine that ecological limits on carrying
207 capacity present the most parsimonious explanation for the latitudinal diversity gradient, though
208 all categories of hypothesis produced gradients. Worm and Tittensor (2018) were able to recreate
209 diversity gradients with simulations using only two parameters: temperature and community size.
210 Also using simulations, Vellend (2005) found that by increasing environmental heterogeneity,

211 rare species increased in abundance so much that the population size and genetic diversity of
212 other species in the community decreased. Our findings, explaining 75 and 90% of the variation
213 in genetic diversity and species richness respectively, provide strong empirical support for the
214 theory-based inferences described above and extend them to the genetic level—multiple eco-
215 evolutionary processes simultaneously produce the gradients in genetic and species level
216 biodiversity we see in nature.
217
218 Communities with diverse species provide important ecosystem functions and services that
219 contribute to human physical and psychological well-being. Ecosystem sustainability in the face
220 of environmental perturbations, occurring more frequently due to human causes, depends on the
221 resiliency of landscapes, communities, and populations (Oliver et al. 2015). The intimate
222 connections between the environment, species richness and genetic diversity we find here
223 suggest that changes on one level can cascade throughout the system and profoundly reshape
224 broad patterns of global biodiversity across multiple biological levels in ways we do not yet fully
225 grasp. Tradeoffs between genetic and species level biodiversity present a conundrum for
226 management practitioners because complex environments that are hotspots for species richness
227 are more likely to harbor relatively low intraspecific genetic variation, and consequently
228 populations that may be less resilient to environmental change. Thus it appears designating
229 conservation areas with the preservation of both species and genetic diversity in mind is not an
230 advisable strategy. Instead, programs focusing on conserving environments and native biological
231 communities should be separate from those aiming to preserve population size and standing
232 genetic variation.
233

Acknowledgements: We would like to thank the Population Ecology and Evolutionary Genetics group for their feedback on this manuscript. We are also grateful to the authors whose work provided the raw data for this synthesis. C.S. and C.J.G. were supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to C.J.G. C.S. was also supported by a U. Manitoba Graduate Fellowship, and a U. Manitoba Graduate Enhancement of Tri-council funding grant to C.J.G.

Author contributions: C.J.G. and C.S. conceptualized the study. C.S., S.D. and C.J.G. designed the study and C.S. conducted the statistical analysis with input from S.D. and C.J.G. All authors contributed to data interpretation. C.S. wrote the first draft of the manuscript and all authors participated in editing subsequent manuscript drafts.

References

Bazin, E., Glémin, S., and Galtier, N. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science. **312**(5773): 570–572. doi:10.1126/science.1122033.

Blanchet, G.F., Legendre, P., and Borcard. 2008. Forward Selection of Explanatory Variables. Ecology **89**(9): 2623–2632.

Borcard, D., and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Modell. **153**(1–2): 51–68. doi:10.1016/S0304-3800(01)00501-4.

Borcard, D., Legendre, P., Avois-Jacquet, C., and Tuomisto, H. 2004. Dissecting the Spatial Structure of Ecological Data at Multiple Scales. Ecology **85**(7): 1826–1832.

Brodie, J.F. 2019. Environmental limits to mammal diversity vary with latitude and global temperature. Ecol. Lett. **22**(3): 480–485. doi:10.1111/ele.13206.

Charlesworth, B., and Charlesworth, D. 2010. Elements of Evolutionary Genetics. Roberts & Company Publishers, Greenwood Village, Colorado, USA.

Currie, D.J. 1991. Energy and Large-Scale Patterns of Animal- and Plant-Species Richness. Am. Nat. **137**(1): 27–49.

Damuth, J. 1981. Population density and body size in mammals. Nature **290**(April): 1980–1981.

Dray, S., Legendre, P., and Peres-Neto, P.R. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Modell. **196**(3–4): 483–493. doi:10.1016/j.ecolmodel.2006.02.015.

Etienne, R.S., Cabral, J.S., Hagen, O., Hartig, F., Hurlbert, A.H., Pellissier, L., Pontarp, M., and Storch, D. 2019. A Minimal Model for the Latitudinal Diversity Gradient Suggests a Dominant Role for Ecological Limits. Am. Nat. **194**(5): E000–E000. doi:10.1086/705243.

Fisher, J.B., Whittaker, R.J., and Malhi, Y. 2011. ET come home: Potential evapotranspiration in geographical ecology. Glob. Ecol. Biogeogr. **20**(1): 1–18. doi:10.1111/j.1466-8238.2010.00578.x.

Hämälä, T., Mattila, T.M., and Savolainen, O. 2018. Local adaptation and ecological differentiation under selection, migration and drift in Arabidopsis lyrata. Evolution. doi:10.1111/evo.13502.

Hubbell, S.P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton NJ.

IUCN. 2019. The IUCN Red List of Threatened Species. Version 2019-1. Available from https://www.iucnredlist.org.

Jiménez-Alfaro, B., Chytrý, M., Mucina, L., Grace, J.B., and Rejmánek, M. 2016. Disentangling vegetation diversity from climate-energy and habitat heterogeneity for explaining animal geographic patterns. Ecol. Evol. **6**(5): 1515–1526. doi:10.1002/ece3.1972.

Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W., Boakes, E.H., Carbone, C., Connolly, C., Cutts, M.J., Foster, J.K., Grenyer, R., Habib, M., Plaster, C.A., Price, S.A., Rigby, E.A., Rist, J., Teacher, A., Bininda-Emonds, O.R.P., Gittleman, J.L., Mace, G.M., and Purvis, A. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology **90**(9): 2648–2648. doi:10.1890/08-1494.1.

Kadmon, R., and Allouche, O. 2007. Integrating the effects of area, isolation, and habitat heterogeneity on species diversity: A unification of island biogeography and niche theory. Am. Nat. **170**(3): 443–454. doi:10.1086/519853.

Kerr, J.T., and Packer, L. 1997. Habitat heterogeneity as a determinant of mammal species richness. Nature **385**: 253–254.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

Kreft, H., and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. Proc. Natl. Acad. Sci. **104**(14): 5925–5930. doi:10.1073/pnas.0608361104.

Lefcheck, J., Byrnes, J., and Grace, J. 2019. piecewiseSEM: Piecewise Structural Equation Modeling. Available from https://cran.r-project.org/package=piecewiseSEM.

Lefcheck, J.S. 2016. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods Ecol. Evol. **7**(5): 573–579. doi:10.1111/2041-210X.12512.

Manel, S., Gugerli, F., Thuiller, W., Alvarez, N., Legendre, P., Holderegger, R., Gielly, L., and Taberlet, P. 2012. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. Mol. Ecol. **21**(15): 3729–3738. doi:10.1111/j.1365-294x.2012.05656.x.

Merckx, T., Souffreau, C., Kaiser, A., Baardsen, L.F., Backeljau, T., Bonte, D., Brans, K.I., Cours, M., Dahirel, M., Debortoli, N., De Wolf, K., Engelen, J.M.T., Fontaneto, D., Gianuca, A.T., Govaert, L., Hendrickx, F., Higuti, J., Lens, L., Martens, K., Matheve, H., Matthysen, E., Piano, E., Sablon, R., Schön, I., Van Doninck, K., De Meester, L., and Van Dyck, H. 2018. Body-size shifts in aquatic and terrestrial urban communities. Nature **558**(7708): 113–116. doi:10.1038/s41586-018-0140-0.

Miraldo, A., Li, S., Borregaard, M.K., Florez-Rodriguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K.A., and Nogues-Bravo, D. 2016. An Anthropocene map of genetic diversity. Science. **353**(6307): 1532–1535. doi:10.1126/science.aaf4381.

Mittell, E.A., Nakagawa, S., and Hadfield, J.D. 2015. Are molecular markers useful predictors of adaptive potential? Ecol. Lett. **18**(8): 772–778. doi:10.1111/ele.12454.

Nabholz, B., Mauffrey, J.F., Bazin, E., Galtier, N., and Glemin, S. 2008. Determination of mitochondrial genetic diversity in mammals. Genetics **178**(1): 351–361. doi:10.1534/genetics.107.073346.

NOAA, and U.S. National Geophysical Data Center. (n.d.). TerrainBase, release 1.0. Boulder CO. Available from https://nelson.wisc.edu/sage/data-and-models/atlas/maps.php?datasetid=28&includerelatedlinks=1&dataset=28.

Oliver, T.H., Heard, M.S., Isaac, N.J.B., Roy, D.B., Procter, D., Eigenbrod, F., Freckleton, R., Hector, A., Orme, C.D.L., Petchey, O.L., Proença, V., Raffaelli, D., Suttle, K.B., Mace, G.M., Martín-López, B., Woodcock, B.A., and Bullock, J.M. 2015. Biodiversity and Resilience of Ecosystem Functions. Trends Ecol. Evol. **30**(11): 673–684. Elsevier Ltd. doi:10.1016/j.tree.2015.08.009.

Pautasso, M., and Gaston, K.J. 2005. Resources and global avian assemblage structure in forests. Ecol. Lett. **8**(3): 282–289. doi:10.1111/j.1461-0248.2005.00724.x.

Pontarp, M., Bunnefeld, L., Cabral, J.S., Etienne, R.S., Fritz, S.A., Gillespie, R., Graham, C.H., Hagen, O., Hartig, F., Huang, S., Jansson, R., Maliet, O., Münkemüller, T., Pellissier, L., Rangel, T.F., Storch, D., Wiegand, T., and Hurlbert, A.H. 2019. The Latitudinal Diversity Gradient: Novel Understanding through Mechanistic Eco-evolutionary Models. Trends Ecol. Evol.: 1–13. doi:10.1016/j.tree.2018.11.009.

Pontarp, M., and Wiens, J.J. 2017. The origin of species richness patterns along environmental gradients: uniting explanations based on time, diversification rate and carrying capacity. J. Biogeogr. **44**(4): 722–735. doi:10.1111/jbi.12896.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria. Available from http://www.r-project.org/.

Schmidt, C., Domaratzki, M., Kinnunen, R.P., Bowman, J., and Garroway, C.J. 2020a. Data from: Continent-wide effects of urbanization on bird and mammal genetic diversity. Dryad

Data Repository. doi:10.5061/dryad.cz8w9gj0c.

Schmidt, C., Domaratzki, M., Kinnunen, R.P., Bowman, J., and Garroway, C.J. 2020b. Continent-wide effects of urbanization on bird and mammal genetic diversity. Proc. R. Soc. B Biol. Sci. **287**(1920): 20192497. doi:10.1098/rspb.2019.2497.

Shipley, B. 2016. Cause and Correlation in Biology. *In* Cause and Correlation in Biology, 2nd edition. Cambridge University Press, Cambridge.

Simpson, G.G. 1964. Species Density of North American Recent Mammals. Syst. Zool. **13**(2): 57–73.

Storch, D., Bohdalková, E., and Okie, J. 2018. The more-individuals hypothesis revisited: the role of community abundance in species richness regulation and the productivity–diversity relationship. Ecol. Lett. **21**(6): 920–937. doi:10.1111/ele.12941.

Theodoridis, S., Fordham, D.A., Nogues-bravo, D., and Brown, S.C. 2020. Evolutionary history and past climate change shape the distribution of genetic diversity in terrestrial mammals. Nat. Commun.: 1–11. Springer US. doi:10.1038/s41467-020-16449-5.

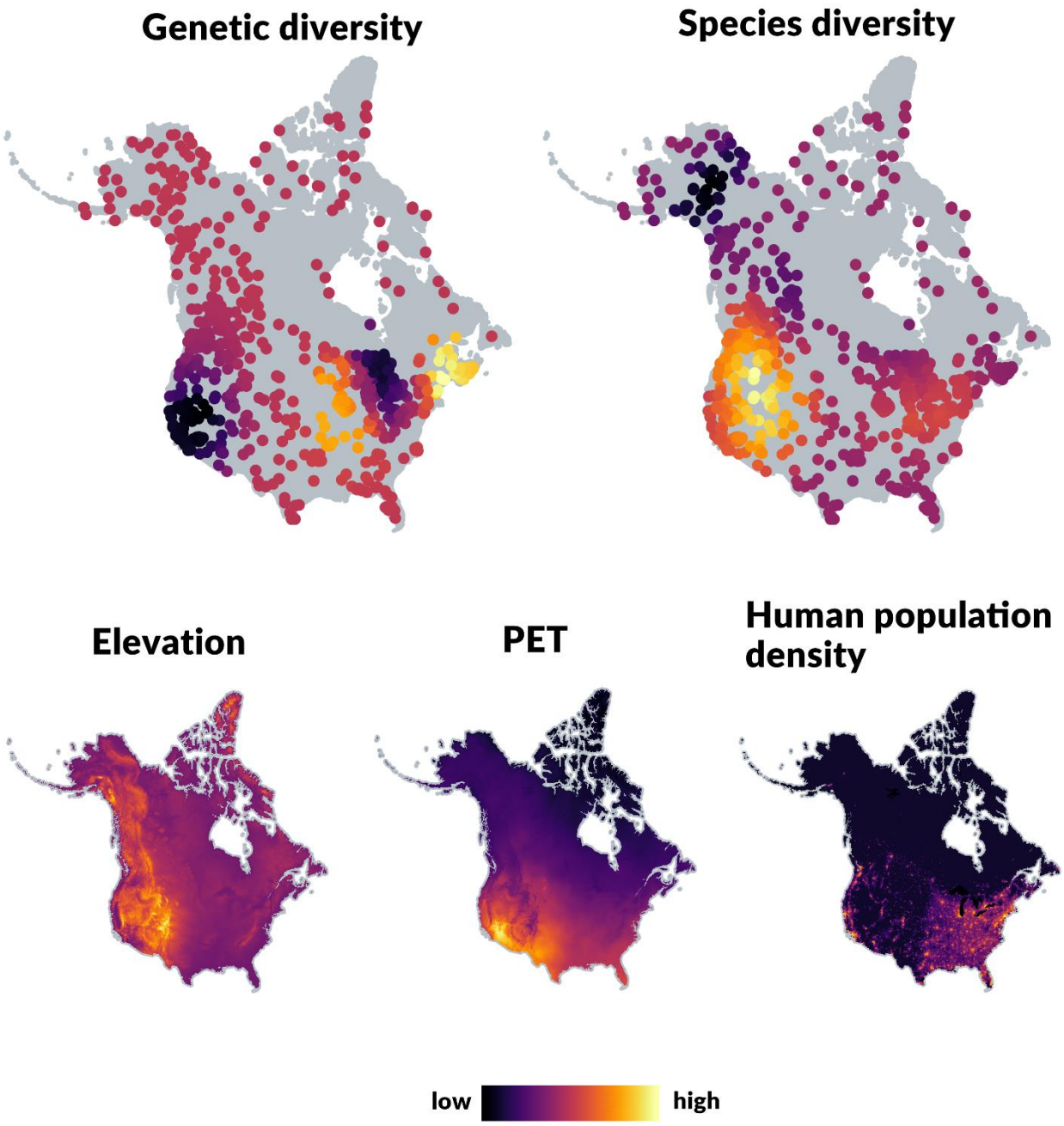
Trabucco, A., and Zomer, R. 2019. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. doi:10.6084/m9.figshare.7504448.v3.

Vellend, M. 2005. Species Diversity and Genetic Diversity: Parallel Processes and Correlated Patterns. Am. Nat. **166**(2): 199–215. doi:10.1086/431318.

Weir, B.S., and Goudet, J. 2017. A Unified Characterization of Population Structure. Genetics **206**: 2085–2103.

Willig, M.R., Kaufman, D.M., and Stevens, R.D. 2003. Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. Annu. Rev. Ecol. Evol. Syst. **34**(1): 273–309. doi:10.1146/annurev.ecolsys.34.012103.144032.

Worm, B., and Tittensor, D.P. 2018. A Theory of Global Biodiversity. Princeton University Press, Princeton, New Jersey.



366

367

368 **Fig. 1.** Maps depicting spatial patterns of biodiversity and environmental factors. *(Top row)*
369 Points are the locations of 801 North American mammal populations for which raw
370 microsatellite data was available in public repositories. Point color indicates predicted values of
371 genetic diversity and species richness based on spatial patterns detected in the data. (*Bottom row*)
372 Maps showing the three environmental variables which we tested for simultaneous effects on
373 genetic diversity and species richness.

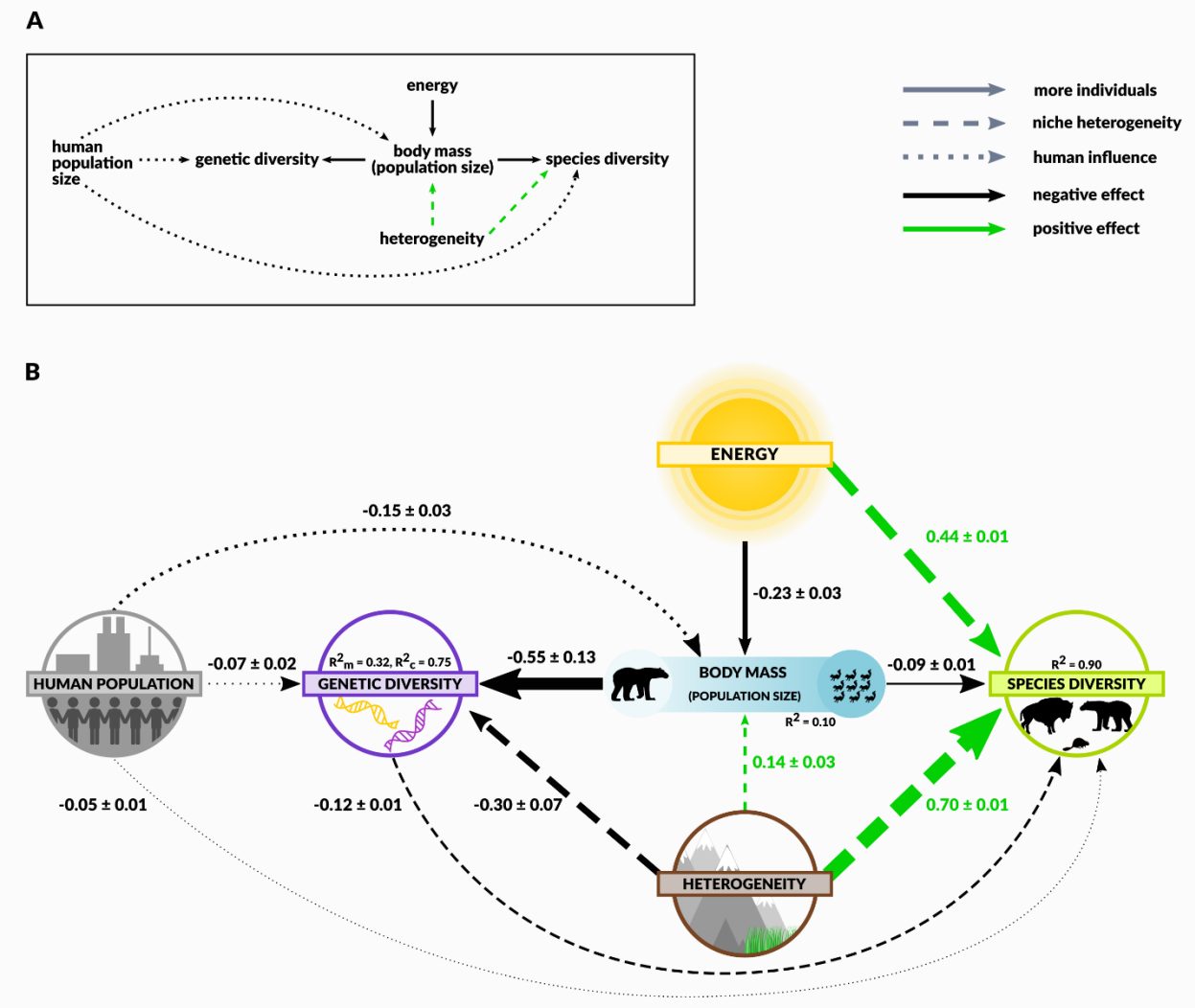


Fig. 2. (a) Our conceptual hypothesis network combining the more individuals hypothesis (solid lines) with the effects of environmental heterogeneity (dashed lines) and human presence (dotted lines). Arrows represent unidirectional relationships between variables. (b) Structural equation model results. Green and black lines positive and negative relationships, respectively. Line widths reflect coefficient estimates, which are listed above each path with standard errors. R^2 values are the amount of variation explained for each response variable. Mass and species richness were measured at the species level, and genetic diversity was measured at the population level and fit with a random effect for species: R^2_m is the variation explained by fixed effects only, and R^2_c is the variation explained by fixed and random effects.

Supplementary Information

Methods

Data assembly

Genetic diversity and body size. We obtained estimates of neutral genetic diversity at 801 sites across North America from the database compiled by Schmidt et al. (Schmidt et al. 2020a, 2020b), where diversity metrics for each site were computed from raw georeferenced microsatellite data . We chose gene diversity at each site as our metric for genetic diversity. Gene diversity estimates the richness and evenness of alleles in a population and is minimally affected by sample size (Charlesworth and Charlesworth 2010). Next, we recorded mean adult body mass (g) for each species using data from the PanTHERIA database (Jones et al. 2009). Mass was log-transformed in our models.

Species richness. We downloaded range maps for terrestrial mammals native to North America from the IUCN Red List database (IUCN 2019). Ranges for all species were mapped in a single Esri shapefile layer with polygon features. We filtered these maps to retain ranges for extant, native, resident, mainland species in ArcMap Desktop 10.3.1 (ESRI, Redlands, CA). We estimated species richness at each of our genetic diversity sample sites to describe spatial variation at both levels of biodiversity across North America. We used a spatial join to count the number of species ranges which overlapped each site to provide a site-level index of species richness. For a broader scale measure of species richness, we also measured richness at the level of each species' range. We did this to avoid introducing bias in biodiversity patterns at the continental scale due to the spread of our genetic diversity sites, which did not consistently sample a species' entire range. We computed species-level species richness by counting the

number of intersecting species ranges using a spatial join in ArcMap. To correct for potential biases due to differences in range size, we divided the number of overlapping ranges by the species' range area (km^2).

Environmental variables. Potential evapotranspiration measures the atmosphere's ability to remove water from the Earth's surface, and is an indicator of atmospheric energy availability. Potential evapotranspiration is one of the strongest environmental correlates of species richness in mammals (Currie 1991; Kreft and Jetz 2007; Fisher et al. 2011; Jiménez-Alfaro et al. 2016). We estimated mean potential evapotranspiration (mm/yr) within each species' range using annual potential evapotranspiration data from 1970-2000 available via the CGIAR Consortium for Spatial Information (Trabucco and Zomer 2019). We used a global topography map (NOAA and U.S. National Geophysical Data Center) to record the range in elevation across focal species ranges to quantify environmental heterogeneity. We also corrected elevation range for potential biases introduced by species range area, because larger ranges tended to encompass greater topographical heterogeneity. Human population sizes were recorded for each site in the aforementioned genetic diversity database (Schmidt et al. 2020a).

Analysis

Spatial patterns in genetic diversity and species richness. All analyses were conducted in R (version 3.6.1, R Core Team 2013). Our first step was to identify spatial patterns in genetic diversity. We accomplished this by adopting a method used in landscape genetics to control for unmeasured environmental variables when investigating adaptive variation associated with environmental factors (Manel et al. 2012). Distance-based Moran's eigenvector maps (MEMs) detect spatial patterns in data from a modified matrix of distances between sites—a 'neighbor'

matrix—whose eigenvalues are proportional to Moran's I index of spatial autocorrelation (Borcard and Legendre 2002; Borcard et al. 2004; Dray et al. 2006). MEMs are vectors that represent spatial relationships between sites at all scales detectable by the sampling scheme, and can be included in linear models to account for effects of unknown spatial processes. Our next step was to determine which MEMs reflected important spatial patterns in genetic diversity and site-level species richness, using a forward selection procedure (Blanchet et al. 2008). This gave us two sets of MEMs which described spatial patterns present in genetic diversity and species richness. To produce continental maps of genetic and species levels of biodiversity we selected MEMs which modeled broad scale spatial patterns based on Moran's I (MEMs with Moran's I > 0.25). We then fit individual linear regression models for species and genetic diversity with corresponding MEMs, and plotted the predicted values.

Variation partitioning. We then examined the extent to which genetic diversity and species richness covary spatially. Because MEMs for species richness and genetic diversity were computed from the same set of coordinates, they were directly comparable: this allowed us to identify shared spatial patterns that might have a common environmental cause. We used linear regressions and variance partitioning to determine what fraction of the total variation in species richness and genetic diversity could be attributed to: (1) non-spatial variation, (2) non-shared shared spatial variation, and (3) shared spatial variation.

Joint analysis of causes of genetic diversity and species richness. Next, we tested the hypothesis that differential carrying capacities are important drivers of biodiversity, specifically the more individuals hypothesis and the effects of environmental heterogeneity and human activity. We developed a causal framework based on these hypotheses (Fig. 3a) which we tested using structural equation modeling (SEM). SEM is an extension of multivariate regression in which a

series of regressions representing causal relationships between variables are assessed as components of a hypothesis network (Shipley 2016). We implemented SEMs in the piecewiseSEM package (Lefcheck 2016; Lefcheck et al. 2019). PiecewiseSEM offers greater flexibility than other SEM software because it uses a local estimation approach where each model is assessed individually (Lefcheck 2016). This frees the user from assuming multivariate normality and linear relationships between variables when using global estimation to evaluate SEM fit. All variables were scaled and centered prior to analysis.

We translated our conceptual diagram (Fig. 3a) into a network of 3 linear models with a single model for each response variable: gene diversity, body size, and species richness. Gene diversity was measured for populations nested within species, thus we used a hierarchical model within the SEM framework to control for species differences by fitting it as a random effect.

Goodness-of-fit in SEM is determined by evaluating whether there are any missing links in the causal structure, i.e. whether adding paths between pairs of variables would be more consistent with the data. In piecewiseSEM missing links are tested using tests of directed separation (Shipley 2016), where the null hypothesis is that the two variables are independent, conditional on other predictors of either variable. Starting with our conceptual model (Fig. 3a), we iteratively updated models by adding links according to tests of directed separation until no further biologically sensible links were suggested. A non-sensible link, for example, would be genetic diversity causing energy availability. We assessed model fit using the *p*-value for the model network, where the null hypothesis is that the model is consistent with the data. Thus, models with $p > 0.05$ are considered acceptable – we fail to reject our causal structure. We also assessed fit using R^2 values for each response variable in the model network. For genetic diversity, we used marginal (R^2_m) and conditional R^2 (R^2_c) values which respectively measure the total

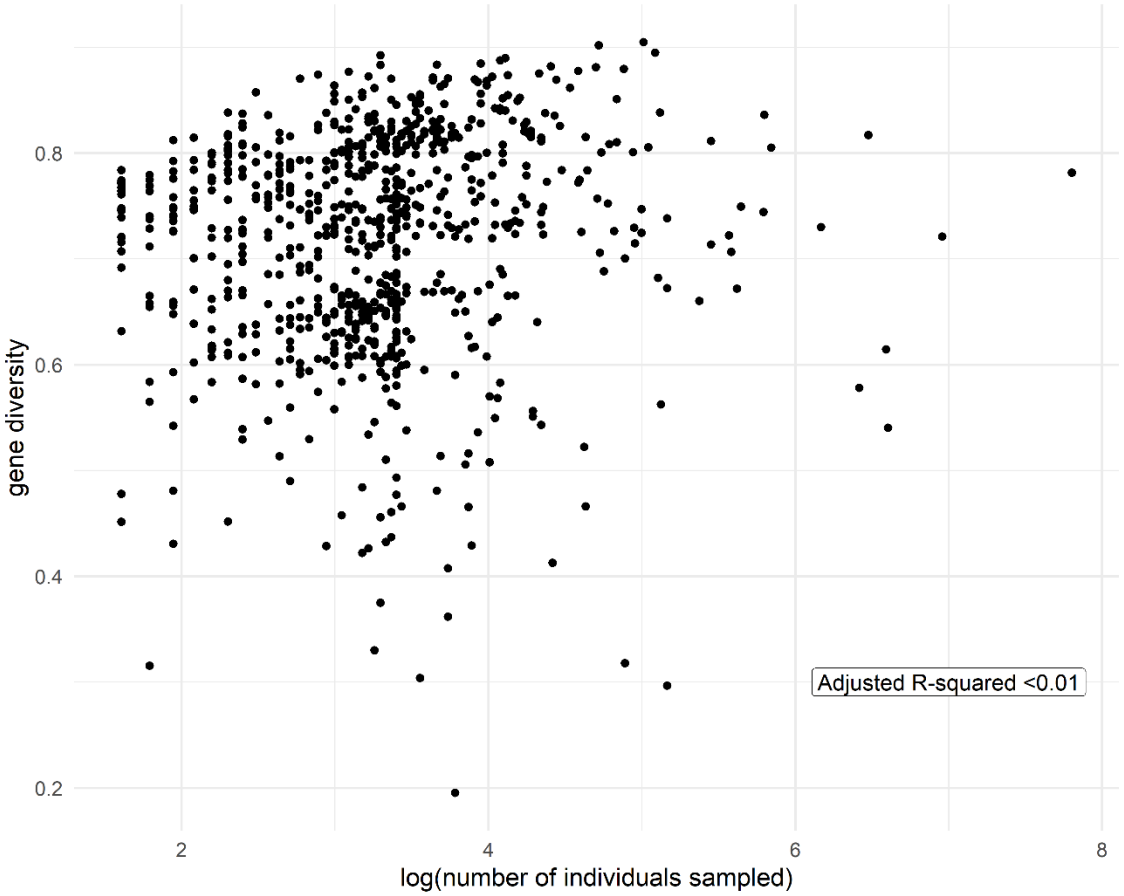
variation explained by fixed effects and the variation explained by both fixed and random effects. We then tested the residuals from component models for spatial autocorrelation using Moran's tests and spatial correlograms. The body size model residuals were not spatially autocorrelated, but genetic diversity and species richness models had statistically significant spatially autocorrelated residuals at very local scales (genetic diversity Moran's I = 0.025, species richness Moran's I = 0.029). These Moran's I values do not indicate strong spatial structure in the data and we decided not to integrate it into our model. The positive spatial autocorrelation at such short distances is likely an artifact of irregular site locations and the hierarchical nature of the data.

Effect of heterogeneity on population divergence. We tested whether topographic heterogeneity caused greater population differentiation using population-specific F_{ST} (Weir and Goudet 2017), a measure of genetic divergence. We controlled for isolation-by-distance by including dbMEMs significantly related to F_{ST} to account for spatial structure. We scaled and centered all variables, then used a linear mixed model of the form: $F_{ST} \sim heterogeneity + MEMs + (1|species)$, controlling for species differences by including it as a random effect.

Supplementary Results

Spatial patterns in genetic diversity and species richness. We detected spatial patterns in genetic diversity at several scales. A total of 199 MEMs corresponded to positive spatial autocorrelation. Of these, 13 explained important spatial variation in gene diversity. In order of increasingly fine spatial scales, significant patterns were MEMs 2, 3, 4, 5, 22, 27, 30, 31, 47, 49, 101, 145, 152. Forty-three MEMs were important predictors of species richness, and 8 of these patterns were shared by genetic diversity (Fig. S1). The cut-off for broad-scale MEMs—defined as those

498 having Moran's I values > 0.25—was MEM 5 for genetic diversity and MEM 11 for species
499 richness.



500

501 **Figure S1.** Plot of gene diversity vs. sample size. Gene diversity as a metric of genetic diversity
502 depends on allele frequencies and is minimally affected by sample size. Larger populations have
503 more rare alleles, which contribute little to gene diversity.

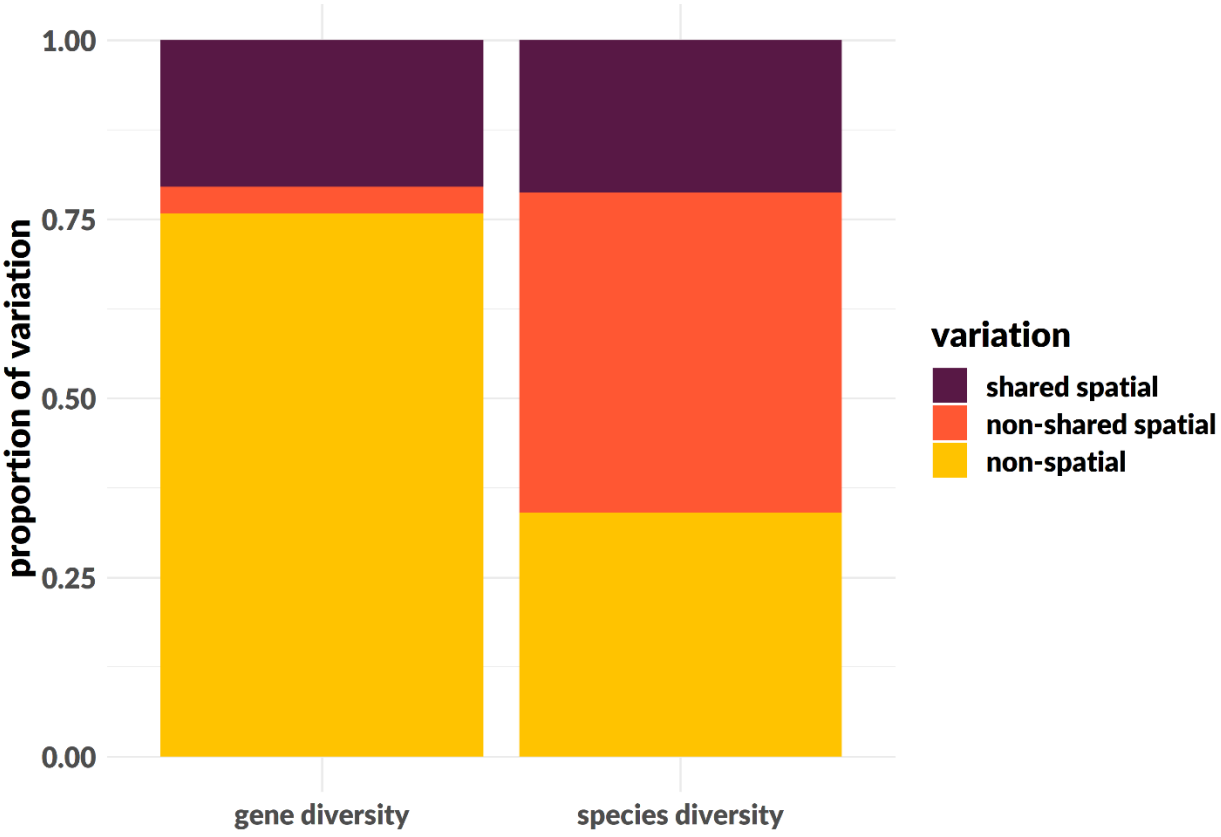


Fig. S2. Variation partitioning results. This graph shows the proportion of variation in genetic diversity and species richness which can be explained by spatial factors, determined using Moran's eigenvector maps (MEMs). Spatial variation is further broken down into shared and non-shared spatial variation. Shared spatial variation is variation in genetic diversity and species richness explained by shared MEMs; non-shared variation is the remaining fraction of spatial variation not accounted for by shared MEMs.

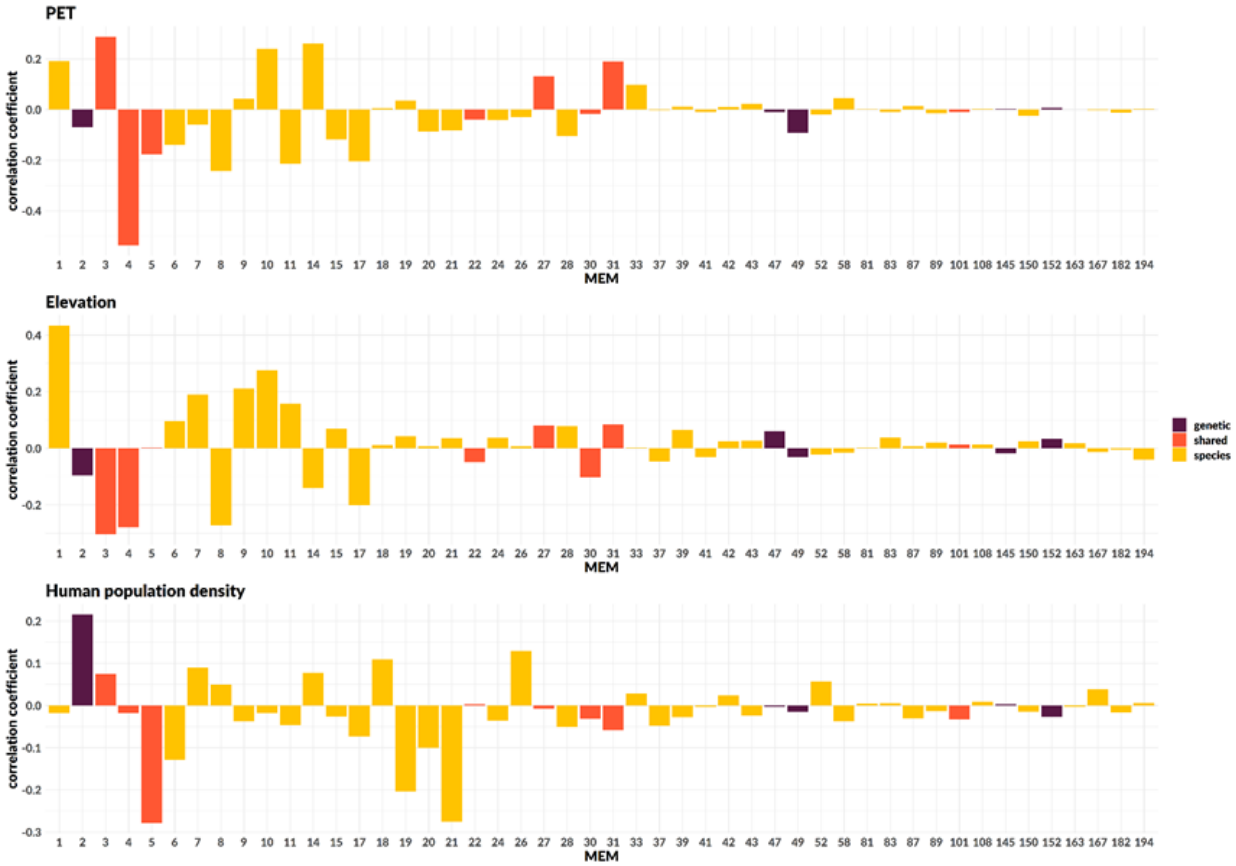


Figure S3. Correlation coefficients for spatial patterns (MEMs) and environmental variables measured at the site level: potential evapotranspiration (PET), elevation, and human population density. MEMs describe spatial patterns in genetic diversity, species richness, or both (shared spatial patterns). MEMs are ordered from broad (MEM1) to fine scale (MEM194) patterns. Strong correlations indicate that environmental variables included in structural equation models account for broad scale spatial patterns present in genetic diversity and species richness.

518 **Table S1.** Path coefficients and standard errors for SEM model (Fisher's C = 2.92, p = 0.23, 2
519 degrees of freedom).

Response	Predictor	Estimate ± SE
genetic diversity	human population density	-0.07 ± 0.02
genetic diversity	mass	-0.55 ± 0.13
genetic diversity	heterogeneity	-0.30 ± 0.07
mass	PET	-0.23 ± 0.03
mass	heterogeneity	0.14 ± 0.03
mass	human population density	-0.15 ± 0.03
species richness	mass	-0.09 ± 0.01
species richness	heterogeneity	0.70 ± 0.01
species richness	human population density	-0.05 ± 0.01
species richness	PET	0.44 ± 0.01
species richness	genetic diversity	-0.12 ± 0.01

520