Title: Continent-wide effects of urbanization on bird and mammal genetic diversity

Authors: C. Schmidt[1*], M. Domaratzki[2], R.P. Kinnunen[1], J. Bowman[3,4], C.J. Garroway[1*]

Affiliations:

[1]Department Biological Sciences, University of Manitoba, Winnipeg, MB, Canada R3T 2N2
[2]Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada R3T 2N2
[3]Environmental and Life Sciences Graduate Program, Trent University, Peterborough, ON, Canada K9L 0G2
[4]Wildlife Research and Monitoring Section, Ontario Ministry of Natural Resources and Forestry, Peterborough, ON, Canada K9J 8M5

***Correspondence to:**
Chloé Schmidt
Biological Sciences Building, University of Manitoba
50 Sifton Rd
Winnipeg, MB, Canada R3T 2N2
Email: schmid46@myumanitoba.ca

Colin Garroway
Biological Sciences Building, University of Manitoba
50 Sifton Rd
Winnipeg, MB, Canada R3T 2N2
Phone: (204) 474-8267; email: colin.garroway@umanitoba.ca

Abstract: Urbanization and associated environmental changes are causing global declines in vertebrate populations. In general, population declines of the magnitudes now detected should lead to reduced effective population sizes for animals living in close proximity to humans. This is cause for concern because effective population sizes set the rate of genetic diversity loss due to genetic drift, the rate of increase in inbreeding, and the efficiency with which selection can spread beneficial alleles. We predicted that the effects of urbanization should decrease effective population size, which would in turn decrease genetic diversity and increase population-level genetic differentiation. To test for such patterns, we repurposed and reanalyzed publicly archived genetic data sets for North American birds and mammals. After filtering, we had usable raw genotypes for 41,023 individuals, sampled from 1,008 locations spanning 41 mammal and 25 bird species. We used census-based urban-rural designations, human population density, and the Human Footprint Index as measures of urbanization and habitat disturbance. As predicted, mammals sampled in more disturbed environments had lower effective population sizes and genetic diversity, and were more genetically differentiated from those in less disturbed environments. There were no consistent effects for birds. This suggests that mammal populations continuing to live in close proximity to humans can generally be expected to have less capacity to respond to further environmental changes, and more likely to suffer from effects of inbreeding.

Keywords: urbanization, genetic diversity, evolution, mammals

Significance statement: The leading cause of contemporary biodiversity loss at the level of populations and species is the wholesale transformation of natural environments by humans. In the span ~50 years vertebrate populations have declined by ~60% on average while the number of threatened and endangered species has increased. These systematic reductions in population size will likely have unintended effects on evolutionary change in animals. Here, we show that environmental degradation consistently erodes the genetic diversity of mammal populations living in close proximity to humans in ways that negatively affect their probability of persistence, compounding direct effects of habitat loss.

Main Text:

Introduction

Human activities are among the most prominent and efficient drivers of contemporary evolution (1). In some cases, human-caused evolution in wild populations is well understood and predictable. For instance, we have a well-founded expectation that populations of pests and disease agents will respond adaptively to our attempts at controlling them (1). It is also clear that humans inadvertently alter evolutionary change in wild populations through land use and habitat degradation (2, 3). Whether the indirect effects of human activities on evolutionary change in species that are not directly targeted by our activities can cause predictable evolutionary outcomes is poorly understood. We hypothesized that urbanization and related land degradation, by limiting population size, could consistently alter the genetic composition of wild populations living near humans in ways that decrease the efficiency of selection and increase inbreeding risk. To test this prediction, we repurposed publicly archived molecular genetic data sets for North

American birds and mammals to test for general relationships between urbanization and the genetic compositions of populations.

Urbanization is one of the most pervasive causes of habitat fragmentation and general landscape change. In addition to the ~700,000 km^2 occupied by cities (4), nearly 75% of the Earth's land surface has been modified by humans, primarily in support of city dwellers (5). This human-caused degradation of the planet's land surface has consistently reduced its capacity to support wildlife (6). As a result, vertebrate populations have on average declined in size by ~60% between 1970 and 2014 (6). Reductions in population size at this level should increase the strength of genetic drift—allele frequency variation due to the random sampling of gametes from one generation to the next. While genetic drift is a neutral evolutionary process that operates independently of the selective value of alleles, it reduces the efficiency of deterministic evolutionary processes like selection by causing allele frequencies to randomly deviate from expected values. When drift is strong relative to selection, random gamete sampling becomes the predominant cause of allele frequency change. In addition, increased drift can eventually lead to reduced mean fitness in populations due to inbreeding depression. If wildlife populations living in close proximity to humans generally experience reductions in population size and connectivity, and thus increased drift, then they may systematically become less genetically diverse than those living in less disturbed environments. By altering a population's genetic composition in this way, human-caused environmental change could make evolutionary responses to such change less efficient.

We tested for general relationships between the human modification of terrestrial habitats and the genetic composition of North American mammals and birds using archived microsatellite data from 41,023 individuals, sampled at 1,008 georeferenced sample sites, spanning 66 species (Table S1, Table S2). In particular, we studied the effects of urbanization and the human footprint (7). Our approach was made possible due to the accumulation of data in public data archives, and a still changing culture of open data in ecological and evolutionary research. Access to raw data originally generated for unrelated purposes allowed for a particularly powerful synthetic analysis: we could consistently calculate population genetic parameters of interest for our question, whether or not they were presented in the original publications, and these calculations are repeatable. In addition, the fact that these data were collected to address different questions reduces the likelihood that study system selection—perhaps a tendency to explore evolutionary responses to humans in systems where such responses are expected—biased our findings.

We developed our specific predictions for the effects of urbanization on wild populations based on basic population genetic theory (Fig. 1). Assuming a finite population of constant size with individuals that randomly mate, die out, and are completely replaced by their offspring each generation, populations will lose genetic diversity at a rate inversely proportional to population size. In reality, natural populations always deviate from these assumptions. Fortunately, we can substitute the concept of effective population size for census population size and the predictive utility of the theory holds. The effective population size is the size of an idealized population which conforms to the assumptions above and produces the same rate of drift as observed in the measured population. We can think of effective population size as a measure of the rate at which

genetic drift causes a population to lose genetic diversity. Nearly all violations of the above assumptions cause the effective population size to be much lower than the census population size, underscoring that drift plays a more important role in determining genetic diversity and the efficiency of selection than what might be expected from census population size alone. Given that urbanization reduces census population size and fragments populations, we predicted that the strength of genetic drift would increase with increasing proximity to environmental disturbances: this would produce smaller effective population sizes, decreased genetic diversity, and increased genetic differentiation.

We chose to analyze data sets that used neutral microsatellite markers because microsatellites were the most common molecular marker type available in data repositories, and because the evolutionary processes that we are interested in are best measured with neutral markers. Although the number of loci surveyed in microsatellite studies is often small relative to surveys of genome-wide markers, the typical number of microsatellites used (~10 loci) in fact estimates genome-wide diversity well with little gain in accuracy with additional genotyping (8). While questions about adaptive genetic variation are interesting, adaptive diversity is currently more difficult to generally define and interpret than neutral genetic diversity, and there are still relatively few data sets suitable for this type of multi-population and multi-species analysis.

We tested for effects of urbanization and the human footprint on estimates of four population genetic parameters calculated for each site (196 bird sites, of which 129 sampled non-migratory species and were reanalyzed separately, and 812 mammal sites, Figure 2; Table S1). We estimated contemporary effective population size of the parental generation using a single

sample linkage disequilibrium method to quantify genetic drift (9–11). Of available methods, this approach is one of the most accurate and is relatively robust to departures from underlying assumptions about population structure (12). Estimators of effective population size perform poorly when sampling error swamps signals of genetic drift, and this meant that effective population size was not estimable at some sites, which were excluded from analysis (see Methods for details). Gene diversity (13) is a measure of genetic diversity that accounts for the evenness and abundance of alleles, and it is not significantly affected by sample size or rare alleles (14). We calculated allelic richness, the number of alleles per locus corrected for sample size, as a second measure of genetic diversity. To quantify genetic differentiation among sites, we estimated site-specific F_{ST} (15).

We focused our analyses on the continental United States and Canada due to the historical and demographic similarities of cities and land-usage in this region (16), and to ensure that species have had broadly similar exposures to past climate variation (17). We chose three indices of urbanization and environmental disturbance. First, we classified a sample as coming from an urban or rural site based on United States Census Bureau (18) and Statistics Canada (19) classifications of urban areas and population centers. Second, we measured human population density at each site, which may capture aspects of the continuous nature of the effects of human presence that would not be apparent in the binary urban-rural classification. Lastly, we used the Human Footprint Index (7) as a measure of disturbance because it incorporates data from multiple land use types including human population density, built-up areas, nighttime lights, land cover, and human access to coastlines, roads, railways, and navigable rivers.

Finally, we had to account for factors other than human disturbance that could affect genetic diversity. Neutral genetic diversity varies with species-level life history traits (20), so when fitting models we allowed relationships to vary among species by treating species as a random effect in a generalized linear mixed modelling framework. Genetic diversity is also affected by regional historical contingencies which would be difficult to specifically identify without detailed knowledge of each species and region in our data set. Such events will, however, produce spatial patterns in our genetic measures. These spatial patterns are detectable and can be controlled for—even if their causes are unknown—using distance-based Moran's Eigenvector Maps (dbMEMs) (21–23). Briefly, dbMEMs are orthogonal spatially explicit eigenvectors that summarize spatial autocorrelation (Moran's I) patterns in data across all scales. We used dbMEMs that described spatial variation in our measures of the genetic composition of sample sites in our regression models to explicitly account for processes causing spatial patterns in the data (22, 24).

To test for relationships between genetic diversity and urbanization, we treated each of our four population genetic parameters (effective population size, allelic richness, gene diversity, and site-specific F_{ST}) as dependent variables in a series of generalized linear mixed models. We first fit each population genetic parameter to each environmental variable (urban-rural, human population density, and Human Footprint) in separate models that also contained terms for species as a random effect, and spatial variables (dbMEMs) when they were important descriptors of spatial patterns in genetic data. Finally, we fit a null model to each population genetic parameter that contained the random effect for species and spatial variables only. We fit these models for bird and mammal data independently. Migratory behavior in birds may affect

spatial patterns in genetic diversity depending on where samples were taken, and whether they were sampled during the breeding season. Therefore, we also ran these models separately for non-migratory birds only (7 species, 129 sites; Table S1).

Results and Discussion

Relationships between all measures of urbanization and the genetic composition of mammal populations were statistically clear, consistent, and as predicted (see the position of parameter estimates and the breadth of 95% confidence intervals in Fig 3a). Effective population size, allelic richness, and gene diversity were each negatively related to the measures of urbanization, and sites sampled in more disturbed areas were the most genetically differentiated (Fig. 3a; Table 1). Contrasting these clear relationships, we found no consistent evidence for effects of urbanization and the human footprint on the genetic composition of non-migratory bird samples when analyzed alone (Fig. 3b; Table 1), or when migratory and non-migratory species were combined for analyses (Fig. S1). When non-migratory birds were considered alone, there was a negative relationship between the effective population size at a site and the site's Human Footprint Index. When considering both migratory and non-migratory birds together, there were positive relationships between allelic richness and the Human Footprint Index, and our binary urban-rural sample site designation. No other relationships within the bird data were detected. We note that (see Methods for details) our data structure allowed us to fit models with random slopes and intercepts for mammal species, whereas we could only fit random intercept models for birds. Random intercept only models suffer from more false positives than random slope and intercept models. Thus, the consistent effects for mammals arise from more conservative model

fitting, while more care is needed when interpreting the statistical relationships detected for birds.

To assess model fits we estimated marginal R^2 (R^2_m), the variance explained by the fixed effects, and conditional R^2 (R^2_c), the variance explained by both fixed and random effects (25). For mammals, all models containing indices of human disturbance explained more variation in the genetic composition of populations than null models (Table 1). Human population density explained the most variation in each measure of the genetic composition of mammal sample sites (effective population size: R^2_m 0.27; R^2_c 0.56; gene diversity R^2_m 0.07; R^2_c 0.81; allelic richness R^2_m 0.06; R^2_c 0.74; F_{ST} R^2_m 0.17; R^2_c 0.44; Table 1). Species level variation among both mammals and birds was important and well-captured by random effects, particularly for gene diversity and allelic richness (Table 1).

The lack of consistent evidence for genetic effects of urbanization on birds could in part be due to differences in movement ability. Cities and their surrounding areas are characterized by disjoint patches of habitat interspersed among paved surfaces, buildings, and grassy or agricultural areas (26). Birds' ability to fly may buffer against the effects of habitat fragmentation and allow for gene flow from undisturbed populations (27) in situations where mammal movements would be more restricted. Indeed, a global analysis of 57 mammal species found that the movements of individuals living in areas with a high Human Footprint Index were considerably reduced relative to those in less disturbed areas (28), which suggests that fragmentation could underlie the patterns we detect. Alternatively, it is also possible that the genetic effects of population declines may simply not yet be detectable for birds. North

American mammal populations have declined more quickly than bird populations (6, 29), and data were available for many fewer bird species. Finally, the life histories of bird species may vary in such ways that we should not expect consistent effects of urbanization across all species. The continued accumulation of data, and more in-depth analyses should resolve these issues.

We currently know very little about the general spatial patterns of intraspecific genetic diversity. There have been some synthetic spatial analyses of raw mitochondrial DNA (mtDNA) sequence variation (30, 31), but to our knowledge there are no other syntheses of spatial patterns in nuclear DNA variation similar to ours. Both Miraldo et al. (30) and Millette et al. (31) reanalyzed raw mtDNA sequence data at global scales across multiple taxa and detected latitudinal patterns of variation in sequence diversity. We detected similar broad spatial patterns in our spatial variables and controlled for them. Both of these groups also explored relationships between measures of human disturbance and sequence diversity, but arrived at contradictory results. Miraldo et al. (30) explored both *cytochrome b* and *cytochrome oxidase subunit I* mtDNA sequence variation in relation to anthropogenic disturbance in mammals and found that *cytochrome oxidase subunit I* genetic variation increased in less disturbed environments, but *cytochrome b* variation was not obviously related to human disturbance. Millette et al.'s more recent work (31) spanned more taxa, including both birds and mammals, and examined variation across spatial scales as well as temporal variation in genetic diversity as measured at *cytochrome c oxidase subunit I*. Interestingly, they detected no overarching trend for a loss of genetic diversity associated with proximity to humans, measured by human population density, and no systematic decline in diversity through time. They found considerable spatial, temporal, and taxonomic variation in diversity trends.

How can we explain our results in light of this previous work? Millette et al. (31) describe the caveats associated with the use of mtDNA sequences well. Both of the mtDNA sequences used in the above described work are coding sequences and so do not evolve under neutrality, though they may often behave in a nearly neutral manner. Nevertheless, this means their variability does not necessarily reflect the consequences of habitat degradation on overall population genetic variability. In addition, mtDNA diversity does not accurately reflect population size (32). Habitat fragmentation and reduced population sizes are hypothesized to be the leading mechanisms causing reduced genetic diversity – if these processes are not captured well by these markers, the trend may be difficult to detect. Finally, the studies by Miraldo et al. (30) and Millette et al. (31) were global in scope. This is certainly a strength of their work, but underlying spatial differences may be harder to detect and control for at this scale. By focusing on North America, we attempted to control for variation in the timing and nature of disturbances that would otherwise be difficult at a global scale. Taken together, it is clear that there are interesting spatial trends in intraspecific genetic variation, and the exploration of their underlying causes warrants further study.

Urbanization and the broader human footprint are leading causes of the current high rates of species and population-level biodiversity losses (33, 34). The population genetic patterns we detected reflect patterns in genome-wide nuclear genetic diversity that are ultimately the result of disturbances at the ecosystem, community, and population levels. The consistent effects for mammals across our three environmental measures of disturbance suggest that this pattern is not confined to just urban spaces – human land use is an issue for the genetic diversity of species in

general. This has considerable importance for understanding the current nature and future consequences of biodiversity loss. While monitoring individuals within populations is the best tool for detecting population trends and assessing risk, such direct monitoring of many species is not logistically possible. Our results suggest that calls for intraspecific genetic monitoring programs are warranted and feasible (35). A relatively small number of genetic markers reflects genome-wide diversity well (8) and is capable of detecting the effects of disturbance. In fact, publicly archiving data with publications, originally intended to ensure data posterity, safe-keeping, and research reproducibility, could be better utilized for this task. Access to raw molecular data sets will continue to increase for the foreseeable future, and can be used for monitoring regardless of the original purpose. However, for this to be a useful component of genetic monitoring, more researchers will need to adhere to the standards and best practices for data sharing to maximize reusability (36, 37). This includes using standardized file and metadata formats that are clearly communicated in data package metadata, and including all relevant methodological information. In the data searches presented here, a majority (192/313 = 61%) of datasets were excluded because they did not meet our study criteria (Data S1). However, an additional 36 datasets were excluded for reasons associated with difficulty accessing or interpreting data (Data S1): for example, not being able to download files (i.e., only metadata was available, or only select datasets were deposited), or unclear methodological detail (i.e., no species designations, delineation between study groups was unclear, or lack of spatial reference). We were able to resolve such issues in many cases by contacting the authors, however this might not always be practical for larger studies and limits the ability to automate the data collection process.

Relative to populations in more natural environments, mammal populations in disturbed environments have a reduced capacity to spread beneficial alleles in response to selection pressures, have reduced genetic diversity which can reduce mean population fitness (38, 39), and are more genetically isolated from natural populations. We are extensively and irreversibly creating environmental change while simultaneously reducing the capacity of some populations to evolve in response. Reducing fragmentation and facilitating population connectivity are therefore key to preserving genetic diversity in mammals. Current estimates suggest that by 2050, just 10% of the planet's surface will be unaltered by humans (6). Land transformation processes are eroding genetic diversity in mammals, compounding direct effects of habitat loss in a way that threatens the long-term existence of populations that persist.

Methods

Microsatellite data compilation

Our dataset was comprised of bird and mammal microsatellite data collected from publicly archived, previously published work (Table S2). To create this dataset, we conducted two systematic searches of online databases (Figure S3). We obtained a list of species names for 859 birds and 450 mammals native to Canada and the United States from the IUCN Red List database which includes all species regardless of Red List status. We then queried the Dryad Digital Repository in February 2018 using a python script with the following search terms: species name (e.g. "*Branta canadensis*"), "microsat*", "single tandem*", "short tandem*", and "str". This search yielded 194 unique data packages associated with papers. A second search was

performed in May 2018, this time querying DataOne.org, a network which provides access to data from multiple repositories such as Dryad, the Knowledge Network for Biocomplexity (KNB), and the United States Geographic Survey (USGS). This search was conducted in R using the dataone package (40), a convenient method of querying the DataOne network. Using identical keywords, 237 unique results were generated, 121 of which overlapped with our first search (Figure S3, Data S1).

All data sets were then individually screened for suitability, ensuring: location (Canada and the United States), taxon (native birds and terrestrial mammals), data type (neutral microsatellite markers), and georeferenced sampling (coordinates, maps, or place names). Studies with other factors which may have influenced genetic diversity (e.g. island sites, genetic rescue, translocation, managed or captive populations) were excluded. In total, data from 85 studies were retained for analysis. In a final step, we assured individual sample sites within datasets adhered to our study criteria, and removed those which did not. We maintained the same sample site delineations as in the original work. Criteria for removal from a dataset included island, managed, or captive populations; sites outside of Canada and the United States; and historical samples (where identified). Sites for which we were unable to extract geographic information were also removed, as well as sites with <5 individuals. Any non-neutral microsatellite markers in the data were removed. Next, unique names were assigned to each site, and all datasets were formatted as either STRUCTURE or GENEPOP files and read into R version 3.4.2 (41) using the adegenet package (42).

Geographic site locations

Geographic coordinates provided by the authors were used when available (Table S2). Where spatial location was available for each individual sampled, coordinates were averaged. If site names were provided (e.g. "Yellowstone National Park") with no coordinate reference, we performed a Google Maps search and noted the resulting coordinates. Where applicable, coordinate information was obtained by searching for site names in the Geographic Names Information System (GNIS) or GeoNames database, as was the case for a few datasets from the USGS. In instances where only maps of sampling sites were available, site coordinates were extracted using a reference map in ArcMap version 10.3.1 (ESRI). When georeferencing map images, if sampling locations indicated regions rather than single points, centroid coordinates served as the site location. Centroid coordinates were also calculated as site location for data accompanied by polygon shapefiles as a spatial reference. All coordinates were recorded using the WGS84 (World Geodetic System 1984) coordinate system in decimal degrees, and transformed from other systems or map projections in ArcMap as needed. Finally, when site locations were offshore (42 sites), points were moved to the nearest terrestrial location using the Generate Near Table tool in ArcGIS. Offshore sites (those located in bodies of water) were moved to avoid generating null values for population density and the Human Footprint Index—both of which are high-resolution terrestrial maps which do not extend far past the coast. In some instances, offshore sites were recorded as thus in the original publication, while other times they were generated during the process of obtaining a single location for a site (e.g. the average location of individual coordinates, or centroid location, was in a body of water). Polar bear sites in the Arctic Archipelago constituted half of all offshore sites, while the remainder were coastal species and species sampled near lakes or oceans.

Genetic diversity estimates

We chose to measure gene diversity (13) and allelic richness for each site as measures of genetic diversity. Gene diversity uses allele frequencies to determine the probability that pairs of alleles drawn at random from a population are different, and accounts for both the number and evenness of alleles. This measure is minimally affected by sample size and rare alleles (14), and thus is convenient to use when sample sizes are variable, as is the case here. Gene diversity was calculated using the adegenet package (42). Allelic richness, the number of alleles per locus, is strongly influenced by sample size and effective population size. To account for differences in sample size, we used rarefaction as employed in the R package hierfstat (43) to standardize allele counts to the minimum sample size (n = 5 individuals) across sites (44). Values were then averaged across loci to obtain a single value per site.

Effective population size estimates

We estimated site-specific contemporary effective population sizes using the linkage disequilibrium method for single samples implemented in the software NeEstimator 2.1 (11). The presence of rare alleles produces an upward bias when estimating effective population size which is especially apparent at small sample sizes (45). We therefore set a conservative exclusion threshold (P_{crit}) of 0.1, meaning estimates are made based only on alleles with frequencies higher than this value, which has been shown to markedly reduce bias (45). Linkage disequilibrium methods work well for estimating effective population sizes in small populations, however are less reliable for large populations (46). An estimate of infinity is returned when sampling error swamps detectable signals of genetic drift—which may be the case if too few individuals or loci were sampled to yield any useful information about effective population size.

In these instances, rather than replacing infinity values with arbitrary large values, we chose to exclude all sites for which we were unable to estimate effective population size.

Population-specific F_{ST}

To estimate levels of population differentiation in relation to human disturbance, we measured population-specific F_{ST} (15). In addition to returning an estimate of population structure for single populations, population-specific F_{ST} values (15) also take into account shared ancestry and inbreeding, which can otherwise confound F_{ST} estimates. Moreover, using a population-specific estimator of structure allows us to make comparisons between populations of different species. Population-specific F_{ST} was calculated in R using hierfstat (43), and values were averaged across loci. Additionally, we note that population-specific F_{ST} can only be calculated for species with two or more sample sites, thus sample size was slightly decreased when this condition was not met (Table 1).

Measures of environmental disturbance and urbanization

Urban-rural classification. Our next step was to define urban habitats in North America. The United States Census Bureau and Statistics Canada provide publicly available maps of urban areas and population centers, respectively (18, 19). According to the US Census Bureau, an urban area is defined as any densely developed territory with at least 2500 inhabitants. Statistics Canada defines a population center as any area with a minimum population of 1000, and a population density of 400 persons or more per square kilometer. We considered these international designations of urbanization to be comparable. Canadian and American urban area maps were downloaded as polygon GIS layers and merged into a single layer. Site coordinates

were transformed from WGS84 to the same projection as the urban area maps (GCS North American 1983) in ArcMap to ensure correct alignment. A spatial join was then performed between sites and the urban area layer in order to classify sample locations as "urban" or "nonurban". The search radius parameter was set to 10 km to encompass the entire urban gradient, and account for sprawl. Periurban landscapes which are adjacent to cities may be less densely inhabited, however often encompass areas highly managed or disturbed by humans including farmland, parks, and golf courses; in larger cities, periurban landscapes may extend up to 10 km away from the city center (47). Thus, any site located in, or within 10 km of, an urban area was considered "urban" for the purposes of this study.

Human population density. Human population density was used as a proxy of urbanization and human effects on the environment. In contrast to our binary urban-rural designation of sample sites, human population reflects the continuous distribution of the effects of human presence, and thus should indicate the intensity of the effects of human activity on genetic diversity. A raster map of global population density per square kilometer was obtained for the most recent available year (2000) from NASA's Center for Near Earth Object Studies (https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SEDAC_POP). Next, the raster map and shapefile containing sites as point features were read into R (package rgdal and raster; Bivand et al. 2017, Hijmans 2017). Mean population density was calculated within a 10 km buffer zone around each site.

Human Footprint Index. The Global Human Footprint Index (7, 50) quantifies human influence on a scale of 0 (most wild) to 100 (most transformed) at a 1 km^2 resolution. It provides a more comprehensive assessment of the effects of humans than urban-rural designations or population density alone because it incorporates data from multiple sources of land use. In particular, it

captures human population density, human land use and infrastructure (built-up areas, nighttime lights, land use, and land cover), and human access (coastlines, roads, railways, and navigable rivers). As with the raster map of population density, the Human Footprint Index was imported to R and values per site (within a 10 km buffer zone) calculated using the same method.

Statistical analysis

We modelled birds and mammals separately because we expected them to respond to human disturbance in fundamentally different ways. Within birds, we further classified each species as migratory or non-migratory using information from species accounts in The Birds of North America (51). We then created a separate data subset comprised of only non-migratory species which was analyzed in parallel. Species with a mix of migratory and resident populations were counted as migratory and excluded, as were species with unknown migratory behavior. Current levels of genetic diversity will reflect many past processes in addition to urbanization and human-caused environmental degradation more generally. Such processes include exposure to Pleistocene glaciations as well as species-specific life history traits, such as body mass and longevity, each of which shape effective population size and thus genetic diversity. Because exposure to past environments (17, 52) and life history trait variation (53) vary spatially, we expect the effects of such processes to create spatial variation in genetic diversity. We can account for such spatial patterns by including variables describing spatial patterns in genetic diversity directly in our models, even when the variables themselves are unmeasured. This can be accomplished with distance-based Moran's Eigenvector Maps, or dbMEMs (21–23). The dbMEM analysis we used (R package adespatial (54)) is a type of eigenanalysis based on principal coordinates analysis which produces a set of spatially explicit variables, dbMEMs, that

quantify spatial trends at multiple scales. Because they are orthogonal, dbMEMs can subsequently be included in regression analyses to explicitly model spatial patterns (54). The eigenvalues of the modified distance matrix generated in the first steps of dbMEM analysis are equivalent to Moran's I coefficients of spatial autocorrelation multiplied by a constant (22, 55). Importantly, only positive eigenvalues are considered because negative eigenvalues generate complex principal coordinate axes (56). dbMEMs therefore correspond to positive values of Moran's I, and can account for positive spatial autocorrelation present in the data. Positive spatial autocorrelation occurs when sites nearer to each other are more similar than sites further away, and violates the assumption of independence in our statistical tests. Before undertaking dbMEM, any linear trends in the response variables were removed. Although dbMEM analysis is capable of detecting linear spatial gradients, dbMEMs used to model such trends then cannot be used to recover other, potentially more interesting spatial patterns (21). dbMEM analyses were run in parallel for measures of genetic diversity (gene diversity and allelic richness), population-specific F_{ST}, and effective population size. We were able to calculate gene diversity and allelic richness for all sites, however, removed sites where effective population size was infinite and sites where population-specific F_{ST} could not be computed. To capitalize on available data, we created subsets for genetic diversity, population-specific F_{ST}, and effective population size, omitting rows where the focal variable(s) had null values. For each taxon we thus had 3 data subsets: one for gene diversity and allelic richness, which included all sites; population-specific F_{ST}; and effective population size (Table 1). To select dbMEMs for inclusion in regression analyses, we used forward selection with a p-value criterion (alpha = 0.05) in the SignifReg package (57).

Testing effects of human disturbance on genetic diversity. To test for the effects of human disturbance on genetic diversity, and to determine whether alternate proxies of urbanization would yield similar results, we constructed four linear mixed models per response variable (effective population size, gene diversity, allelic richness, and population-specific F_{ST}). Three of these models included spatial dbMEMs and a measure of disturbance as explanatory variables: (1) urban-rural category, (2) human population density, and (3) Human Footprint Index. The fourth model consisted of dbMEMs only, or, where no dbMEMs were significant, was a null model (Table 1).

Species was included as a random effect in all models to account for species-level variation in genetic diversity, effective population size, and population-specific F_{ST}. Where possible, the random species effect also accommodated potential variation in the level of species' responses to human-caused environmental degradation (random slope models). The random species effect was estimated using restricted maximum likelihood (REML). Random effects account for non-independence of samples within groups and increase the accuracy of parameter estimation (58). Allowing for variation between group means (random intercepts) and effect sizes (random slopes) additionally reduces Type I and Type II error rates (59). However, fitting both random intercepts and slopes requires large sample sizes for groups, and for samples within groups. Among our datasets this was possible for mammals, for which we had substantially more data than birds. We therefore fit all mammal models with random slopes and intercepts, and bird models with both where possible, or else with random intercepts only (Table 1). Random intercept models do not reduce Type I and II error rates to the same extent as models with both random intercepts and slopes, however provide better estimates than models without random effects at all (59).

To evaluate and compare model fits, we computed marginal and conditional R^2 values following the method of Nakagawa and Schielzeth (25) for mixed models using the R package MuMIn (60). Marginal R^2 represents the variance explained by fixed factors, while conditional R^2 is the variance explained by both fixed and random factors (Table 1).

References

1. Palumbi SR (2001) Humans as the World's Greatest Evolutionary Force. *Science* 293(5536):1786–1790.
2. Johnson MTJ, Munshi-South J (2017) Evolution of life in urban environments. *Science* 358(6363):1–11.
3. Otto SP (2018) Adaptation, speciation and extinction in the Anthropocene. *Proc R Soc B* 285(1891):20182047.
4. Liu X, et al. (2018) High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sens Environ* 209:227–239.
5. Barnosky AD, et al. (2012) Approaching a state shift in Earth's biosphere. *Nature* 486(7401):52–58.
6. WWF (2018) *Living Planet Report - 2018: Aiming Higher* eds Grooten M, Almond REA (WWF, Gland, Switzerland) Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:LIVING+PLANET+REPORT+2004#0.
7. Sanderson EW, et al. (2002) The Human Footprint and the Last of the Wild. *Bioscience* 52(10):891.
8. Mittell EA, Nakagawa S, Hadfield JD (2015) Are molecular markers useful predictors of adaptive potential? *Ecol Lett* 18(8):772–778.
9. Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv Genet* 7(2):167–184.
10. Hill WG (1981) Estimation of effective population size from data on linkage

disequilibrium. *Genet Res (Camb)* 38:209–216.

11. Do C, et al. (2014) NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol Ecol Resour* 14(1):209–214.
12. Gilbert KJ, Whitlock MC (2015) Evaluating methods for estimating local effective population size with and without migration. *Evolution* 69(8):2154–2166.
13. Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A* 70(12):3321–3323.
14. Charlesworth B, Charlesworth D (2010) *Elements of Evolutionary Genetics* (Roberts & Company, Greenwood Village).
15. Weir BS, Goudet J (2017) A Unified Characterization of Population Structure. *Genetics* 206:2085–2103.
16. La Sorte FA, McKinney ML, Pyšek P (2007) Compositional similarity among urban floras within and across continents: Biogeographical consequences of human-mediated biotic interchange. *Glob Chang Biol* 13(4):913–921.
17. Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405(6789):907–913.
18. U.S. Census Bureau (2016) TIGER/Line Shapefiles: Urban Areas (machinereadable data files). Available at: https://www.census.gov/geo/maps-data/data/tiger-line.html.
19. Statistics Canada (2016) Population Centres Census Division Cartographic Boundary File. 2016 Census. Statistics Canada Catalogue no. 92-160-X. Available at: https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm.

20. Romiguier J, et al. (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515(7526):261–263.

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

22. Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Modell* 196(3–4):483–493.

23. Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Mol Ecol* 19(17):3824–3835.

24. Dray S, et al. (2012) Community ecology in the age of multivariate multiscale spatial analysis. *Ecol Monogr* 82(3):257–275.

25. Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol Evol* 4(2):133–142.

26. Marzluff JM, Ewing K (2001) Restoration of fragmented landscapes for the conservation of birds: A general framework and specific recommendations for urbanizing landscapes. *Restor Ecol* 9(3):280–292.

27. Buchmann CM, Schurr FM, Nathan R, Jeltsch F (2013) Habitat loss and fragmentation affecting mammal and bird communities: The role of interspecific competition and individual space use. *Ecol Inform* 14:90–98.

28. Tucker MA, et al. (2018) Moving in the Anthropocene: Global reductions in terrestrial mammalian movements. *Science* 359(6374):466–469.

29. Hoffmann M, et al. (2010) The Impact of Conservation on the Status of the World's

Vertebrates. *Science* 330(6010):1503–1509.

30. Miraldo A, et al. (2016) An Anthropocene map of genetic diversity. *Science* 353(6307):1532–1535.

31. Millette KL, et al. (2019) Worldwide impacts of humans on animal genetic diversity. *bioRxiv*:1–24.

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

33. Aronson MFJ, et al. (2014) A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proc R Soc B Biol Sci* 281(20133330). doi:10.1098/rspb.2013.3330.

34. McKinney ML (2006) Urbanization as a major cause of biotic homogenization. *Biol Conserv* 127(3):247–260.

35. Mimura M, et al. (2017) Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evol Appl* 10(2):121–139.

36. Whitlock MC (2011) Data archiving in ecology and evolution: Best practices. *Trends Ecol Evol* 26(2):61–65.

37. Michener WK (2015) Ecological data sharing. *Ecol Inform* 29:33–44.

38. Reed DH, Frankham R (2003) Correlation between Fitness and Genetic Diversity. *Conserv Biol* 17(1):230–237.

39. Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: A time for reappraisal. *Evolution* 64(5):1202–1217.

40. Jones MB, et al. (2017) dataone: R Interface to the DataONE REST API. Available at: https://cran.r-project.org/package=dataone.

41. R Core Team (2013) R: A Language and Environment for Statistical Computing. Available at: http://www.r-project.org/.

42. Jombart T, et al. (2017) adegenet: Exploratory Analysis of Genetic and Genomic Data. Available at: https://cran.r-project.org/package=adegenet.

43. Goudet J, Jombart T (2015) hierfstat: Estimation and Tests of Hierarchical F-Statistics. Available at: https://cran.r-project.org/package=hierfstat.

44. Leberg PL (2002) Estimating allelic richness: Effects of sample size and bottlenecks. *Mol Ecol* 11:2445–2449.

45. Waples RS, Do C (2008) LDNE: A program for estimating effective population size from data on linkage disequilibrium. *Mol Ecol Resour* 8(4):753–756.

46. Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol Appl* 3(3):244–262.

47. Clergeau P, Jokimäki J, Savard JPL (2001) Are urban bird communities influenced by the bird diversity of adjacent landscapes? *J Appl Ecol* 38(5):1122–1134.

48. Bivand R, Keitt T, Rowlingson B (2017) rgdal: Bindings for the Geospatial Data Abstraction Library. Available at: https://cran.r-project.org/package=rgdal.

49. Hijmans RJ (2017) raster: Geographic Data Analysis and Modeling. Available at: https://cran.r-project.org/package=raster.

50. Wildlife Conservation Society - WCS, Center for International Earth Science Information Network - CIESIN - Columbia University (2005) Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic). doi:10.7927/H4M61H5F.

51. Rodewald P ed. (2015) *The Birds of North America* (Cornell Laboratory of Ornithology,

Ithaca, NY). August Available at: https://birdsna-org.uml.idm.oclc.org.

52. Shafer ABA, Cullingham CI, Côté SD, Coltman DW (2010) Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Mol Ecol* 19(21):4589–4621.

53. Stearns SC (1992) *The Evolution of Life Histories* (Oxford University Press, New York).

54. Dray S, et al. (2017) adespatial: Multivariate Multiscale Spatial Analysis. Available at: https://cran.r-project.org/package=adespatial.

55. Moran PAP (1948) The Interpretation of Statistical Maps. *J R Stat Soc Ser B* 10(2):243–251.

56. Borcard D, 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.

57. Kim J, Zambom AZ (2017) SignifReg: Significant Variable Selection in Linear Regression. Available at: https://cran.r-project.org/package=SignifReg.

58. Harrison XA, et al. (2018) A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:e4794.

59. Grueber CE, Nakagawa S, Laws RJ, Jamieson IG (2011) Multimodel inference in ecology and evolution: Challenges and solutions. *J Evol Biol* 24(4):699–711.

60. Barton K (2018) MuMIn: Multi-Model Inference. Available at: https://cran.r-project.org/package=MuMIn.

Figure and table legends

Figure 1. Urbanization is expected to cause smaller effective population sizes, lower genetic diversity, and increased population differentiation in comparison to natural habitats (a). As habitats become increasingly urbanized, they experience greater fragmentation (b), resulting in smaller patch sizes with lower connectivity. Smaller patches limit supportable population sizes wherein genetic drift becomes the predominant evolutionary force and movement between patches in urbanized areas (black circles) becomes difficult, reducing gene flow.

Figure 2. Map of 1,008 sample sites for the 66 mammal and bird species native to North America examined in this study. 812 sites were mammals (black points) and 129 birds (white points). Using microsatellite markers, we calculated effective population size, gene diversity, allelic richness, and population-specific FST for each site.

Figure 3. GLMM coefficients for fixed urban and disturbance effects in mammals (top), and non-migratory birds (bottom; for bird results including migratory species, refer to SI Fig. 1 and SI Table 2). Open circles represent coefficient estimates, bold lines are 90% confidence intervals, and narrow lines are 95% confidence intervals. Sample size differed between variables, i.e. sites where effective population size was not calculable were excluded, and calculation of population-specific FST for all sites within a study required at least two sample sites. Mammals: effective population size n = 639; gene diversity and allelic richness n = 812; FST n = 795. Birds (non-migratory): effective population size n = 87; gene diversity and allelic richness n = 129; FST n = 128.

Table 1. Model summaries for mammals, non-migratory birds, and all birds. Four models were constructed per response variable, each including one of three proxies of urbanization: urban-rural category, human population density (popden), and Human Footprint Index (HFI). The fourth model did not include any measure of urbanization and had only dbMEMs as fixed effects (spatial model), or, where no dbMEMs were selected, a null model. Coefficient of variation, R^2, values are an indicator of model fit; marginal R^2 describes the proportion of variation explained by fixed effects, while conditional R^2 is the variation explained by both fixed and random effects. Random effects are specified in the R package 'lme4' format, where variables preceding the vertical bar (|) indicate random slopes, and two vertical bars (||) indicate uncorrelated random slopes and intercepts.

Figure 1

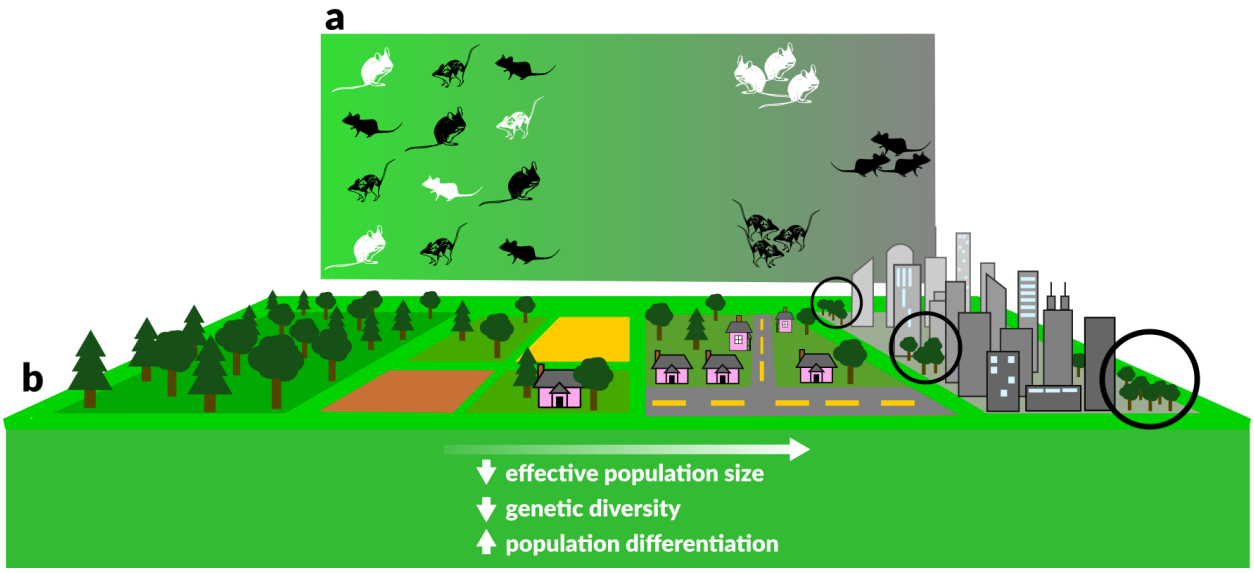

Figure 2

Figure 3

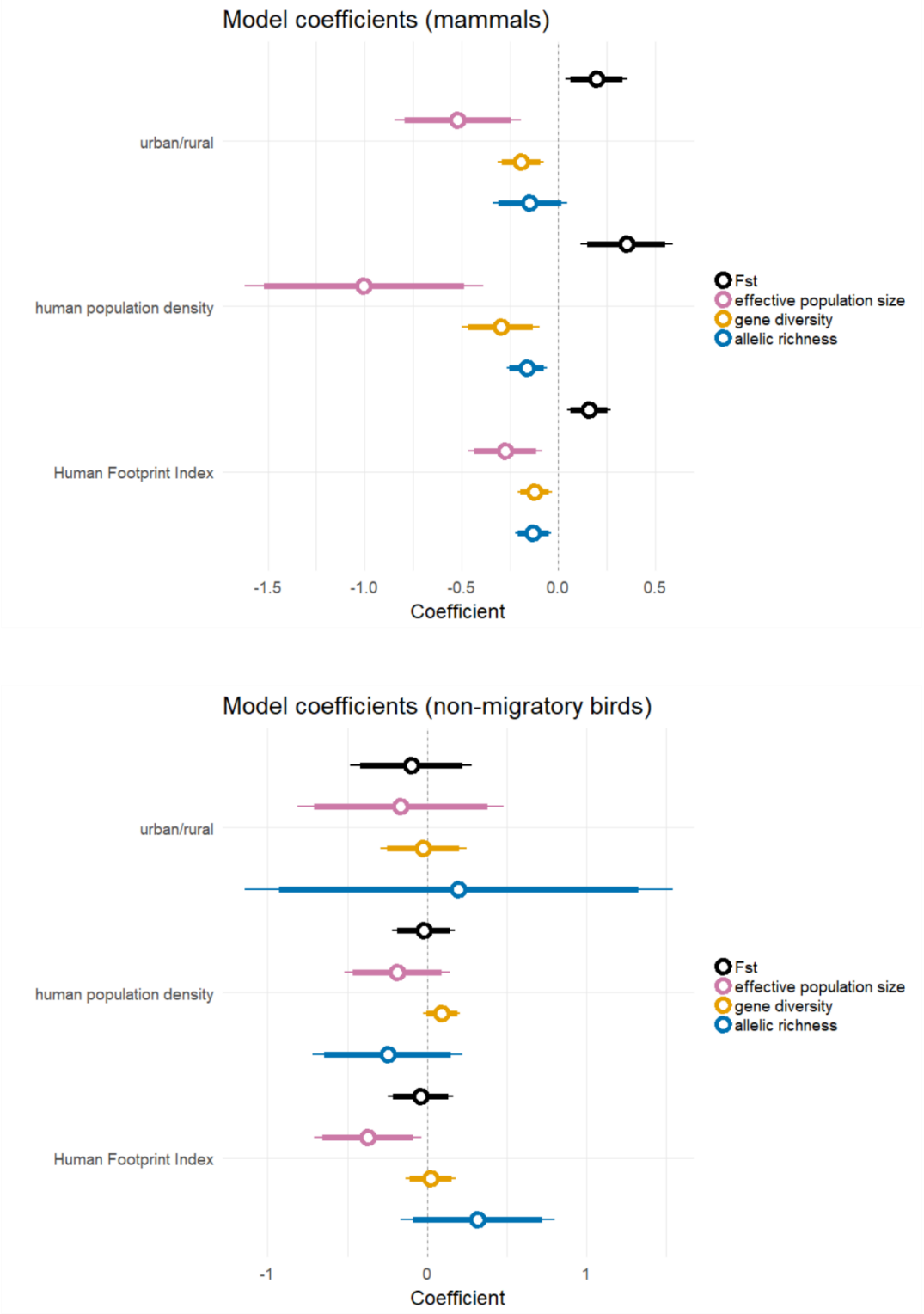

Table 1

Class	variable	sites	Fixed effects		coefficient	95% CI		Marginal R^2	Conditional R^2
			dbMEMS	covariate		lower	upper		
Mammals	effective population size	639	5						
				urban-rural	-0.52	-0.85	-0.19	0.03	0.25
				popden	-1.00	-1.62	-0.39	0.27	0.56
				HFI	-0.27	-0.46	-0.08	0.04	0.27
				none	--	--	--	0.00	0.24
	gene diversity	812	13						
				urban-rural	-0.19	-0.31	-0.07	0.02	0.79
				popden	-0.30	-0.50	-0.10	0.07	0.81
				HFI	-0.12	-0.21	-0.03	0.02	0.80
				none	--	--	--	0.02	0.79
	allelic richness	812	21						
				urban-rural	-0.15	-0.34	0.04	0.05	0.73
				popden	-0.16	-0.27	-0.06	0.06	0.74
				HFI	-0.13	-0.22	-0.04	0.05	0.75
				none	--	--	--	0.05	0.73
	F_{ST}	795	10						
				urban-rural	0.20	0.04	0.36	0.08	0.34
				popden	0.35	0.11	0.59	0.17	0.44
				HFI	0.16	0.05	0.27	0.10	0.37
				none	--	--	--	0.08	0.33
Birds (non-migratory)	effective population size	87	0						
				urban-rural	-0.17	-0.81	0.48	0.00	0.21
				popden	-0.19	-0.52	0.15	0.02	0.17
				HFI	-0.37	-0.71	-0.04	0.07	0.20

				none	--	--	--	0.00	0.21
	gene diversity	129	3						
				urban-rural	-0.02	-0.30	0.25	0.01	0.88
				popden	0.09	-0.02	0.21	0.01	0.89
				HFI	0.02	-0.13	0.18	0.01	0.88
				none	--	--	--	0.01	0.88
	allelic richness	129	0						
				urban-rural	0.20	-1.14	1.54	0.00	0.28
				popden	-0.25	-0.72	0.22	0.01	0.20
				HFI	0.32	-0.17	0.80	0.02	0.30
				none	--	--	--	0.00	0.23
	F_{ST}	128	2						
				urban-rural	-0.10	-0.48	0.28	0.04	0.12
				popden	-0.02	-0.22	0.17	0.04	0.11
				HFI	-0.04	-0.24	0.16	0.04	0.12
				none	--	--	--	0.04	0.11
Birds (all)	effective population size	125	1						
				urban-rural	-0.20	-0.75	0.34	0.06	0.13
				popden	-0.21	-0.48	0.05	0.09	0.12
				HFI	-0.19	-0.46	0.08	0.08	0.14
				none	--	--	--	0.06	0.13
	gene diversity	196	2						
				urban-rural	-0.04	-0.16	0.08	0.00	0.93
				popden	0.04	-0.03	0.10	0.00	0.94
				HFI	-0.01	-0.08	0.05	0.00	0.93
				none	--	--	--	0.00	0.93
	allelic richness	196	1						
				urban-rural	0.84	0.20	1.47	0.04	0.32
				popden	-0.10	-0.43	0.23	0.01	0.30

			HFI	0.35	0.01	0.69	0.04	0.33
			none	--	--	--	0.01	0.30
F_{ST}	190	1						
			urban-rural	0.01	-0.29	0.31	0.01	0.03
			popden	0.03	-0.11	0.18	0.01	0.03
			HFI	0.07	-0.08	0.21	0.02	0.03
			none	--	--	--	0.01	0.03