Title: Worldwide impacts of humans on animal genetic diversity

Running head: Human impacts on genetic diversity

Authors: *Millette K.L[1,5], *Fugère V.[1,2,5], Debyser C.[1,5], Greiner A.[3], Chain F.J.J.[4] and Gonzalez A[1].

[1]McGill University, Department of Biology
[2]Université du Québec à Montréal, Département des Sciences Biologiques
[3]University of Toronto, Department of Ecology & Evolutionary Biology
[4]University of Massachusetts at Lowell, Department of Biological Sciences
[5]These authors contributed equally

*Corresponding authors: katie.millette@mail.mcgill.ca; vincent.fugere@mail.mcgill.ca

Summary

Human impacts on genetic diversity are poorly understood yet critical to understanding the evolutionary capacity of the world's biodiversity. We used global maps of land use and human density to assess human impacts on the intraspecific genetic diversity of 15,946 species of birds, fishes, insects, and mammals over time and across four spatial scales worldwide. We analyzed 164,518 mitochondrial cytochrome *c* oxidase subunit I (COI) sequences to quantify changes in genetic diversity between 1980-2016. We found temporal change in genetic diversity, with both increasing and decreasing trends observed. The magnitude and sign of human impacts on genetic diversity depended on scale and taxon. In contrast, latitude was a strong predictor of diversity in fish, insects, and mammals. Our analyses provide a first worldwide picture of human impacts on animal genetic diversity. A global effort to systematically monitor genetic diversity is needed to fill the gaps in taxonomic and geographic coverage in this dataset.

Introduction

Intraspecific genetic diversity, a measure of the genetic variation within populations, is a fundamental dimension of biodiversity. Intraspecific genetic diversity is a reflection of both past and current evolutionary bottlenecks, as well as an indicator of a population's potential for adaptation to future stressors[1–4]. Understanding the drivers of genetic diversity change worldwide, across taxonomic groups, is of great interest to ecologists and conservation biologists[5–8]. Humans are now acting as an evolutionary force, modifying rates of extinction and colonization, but also altering the intraspecific genetic diversity of plants and animals around the world[9–12]. To our knowledge no global assessment of temporal trends in genetic diversity has been conducted to date, nor have human impacts on such trends being quantified.

Theory predicts that human activities can affect intraspecific genetic diversity via demographic and evolutionary mechanisms[13–15]. Depending on how human disturbances alter selection, drift, gene flow, and mutation rates, intraspecific genetic diversity may decrease, increase, or remain unchanged over time[16]. For example, disturbances like habitat fragmentation and excess harvesting can reduce diversity due to sustained selection, decreased gene flow linked to population isolation, and chronic inbreeding associated with reduced population sizes[17]. Alternatively, human disturbances can maintain or increase genetic diversity through time, for example by magnifying temporal variation in selection, increasing mutation rates (e.g. mutagenic pollutants), or creating environments which favour hybridization and heterozygote advantage[18–21]. Over time and across geographic space, these outcomes can accumulate within populations such that intraspecific genetic diversity reflects a complex combination of past and present evolutionary processes that we are only beginning to investigate at the global scale.

Trends in intraspecific genetic diversity are expected to be scale-dependent as are trends in other dimensions of biodiversity[12,22,23]. Humans impacts should be strongest and most visible at the scales at which they operate, namely that of individual populations. Analyses at larger spatial scales aggregate distant populations which are potentially genetically differentiated, found in heterogeneous habitats, and exposed to varying levels of human impacts, all of which could obfuscate diversity trends. At regional scales, genetic diversity may in fact be highest in human-dominated environments because humans usually settle in areas of high biodiversity[24]. A recent assessment of the global distribution of intraspecific genetic diversity of amphibians and mammals found evidence of reduced genetic diversity in human-impacted regions[25], but the analysis focused on one broad (4° equal area grid cell size) spatial scale of analysis that is

unlikely to reflect population-level processes of all species. Moreover, the lack of time series data prevented the characterization of genetic diversity trends over time.

Here, we report the first large-scale assessment of temporal trends in intraspecific genetic diversity across the world. We overlay a large number of time-referenced mitochondrial cytochrome *c* oxidase subunit I (COI) sequence data for four animal classes (birds, inland and coastal bony fishes, insects, and mammals) on worldwide estimates of human density and land use. We calculated the mean pairwise dissimilarity among all sequences from a population to estimate nucleotide diversity ($\hat{\pi}$), a measure of population genetic diversity. To evaluate the scale-dependence of genetic diversity trends and human impacts, we calculated land use and genetic diversity at four spatial scales, namely 5' (0.08°), 1°, 2°, and 4° square grid cells. We mapped the distribution of sequence data and genetic diversity estimates in time, in geographical space, and across major human impact gradients, and we quantified the drivers of diversity at all spatial scales using a number of spatial and time-series analyses.

Results

Distribution of sequences and genetic diversity

At the finest spatial scale of analysis (5' grid cells or approximately 85 km^2 at the equator), our dataset includes a total of 146,092 COI sequences sampled from 13,936 species of birds (Aves), inland and coastal bony fishes (Actinopterygii), insects (Insecta), and mammals (Mammalia; Table 1). The aggregation of species-specific sequences sampled in the same grid cell at this resolution resulted in a total of 29,436 'populations'. Mapping the geographic location of sequences revealed a spatially heterogeneous pattern of sampling, with 69.5% of sequences originating from North America and Europe (Fig 1a). The number of sequences and species per grid cell were strongly correlated (Fig. S1a). With respect to taxonomy, the dataset was dominated by insect sequences (Fig. 1b; Table 1). Moreover, for all classes, 1-3 speciose orders contributed a large proportion of sequences (Fig. 1b). Only a small proportion of the global number of species of birds, bony fishes, insects, and mammals was represented in our dataset, but this number increased significantly for higher taxonomic levels (families and orders), suggesting a phylogenetically-broad pattern of sampling (Table 1).

The number of COI sequences collected on any given year and deposited on GenBank or BOLD databases has increased for birds, fishes, and insects from 1980 to 2010 (Fig. 1c). In

contrast, mammal sequence collection seems to have reached a peak and remained stable from approximately 1985 to 2005 (Fig 1c). All groups demonstrate a recent (~5 year) decline in the number of sequences, perhaps as a result of the lag between sequence collection and sequence availability in the databases (Fig. 1c). The temporal distribution of species and population number in the database closely matched the temporal distribution of sequence number (Fig. S1b,c). Finally, sequences in the database originate from grid cells with generally higher human population density and more extensive land use than the mean of all grid cells across the world (Fig. 1d,e). Taxonomic and spatial biases are discussed in more detail in Supporting Information ('Supplementary Results').

Global maps of mean population genetic diversity at the 4° scale of analysis suggested higher diversity at lower than higher latitudes (Fig. 2). Potential hotspots of genetic diversity include the Amazon basin for inland fishes, and Madagascar for insects (Fig. 2). When averaged across all populations, species, and grid cells, mean genetic diversity of birds, fishes, and insects showed an increasing trend since 1985, although there was large intra-year variation across much of these spatially-averaged time series (Fig. 3a). Relating genetic diversity to grid cell latitude (absolute values) confirmed the presence of clear latitudinal gradients (Fig. 3b). Finally, at the 4° scale of analysis, no linear relationship was apparent between genetic diversity and longitude, land use intensity, or human population density (Fig. 3c-e). However, genetic diversity of birds, insects, and mammals seemed to vary non-linearly with human population density, and grid cells with large human populations were characterized by very low genetic diversity of insects and mammals (Fig. 3e).

Drivers of intraspecific genetic diversity

Generalized linear mixed models fitted on diversity estimates calculated at four spatial scales confirmed that intraspecific genetic diversity varied strongly with latitude at all spatial scales, for all animal classes except birds (Fig 4; Table S1). Year of sampling had a positive effect on intraspecific genetic diversity in birds at the 5' and 1° scales, and in fish at all spatial scales; however, year had a negative effect on genetic diversity in insects at the 2° scale, and no effect on mammals (Fig. 4; Table S1). Longitude and human population density had no effect on intraspecific genetic diversity across all classes and spatial scales, whereas land use intensity only had a negative effect on genetic diversity in birds at the 5' scale. Negative interaction effects were detected between human density and land use intensity for birds at the 1° scale, and

124 for insects at the 2° and 4° scales. Moreover, there was a negative effect of the interaction
125 between human density and year on genetic diversity for birds (4° scale), fishes (1°, 2° scales),
126 and insects (2° scale). We also found significant interactions between land use intensity and year
127 on genetic diversity, with a positive effect for fish at the 2° scale and a negative effect for
128 mammals at the 5' scale (Fig. 4; Table S1).
129 Additional significant interactions were identified between latitude and human density.
130 Although this interaction had a positive effect on mammal intraspecific genetic diversity at the 5'
131 and 2° scales, the interaction between latitude and land use intensity had a negative effect on
132 insect genetic diversity at the 2° scale. The interaction between latitude and year had an overall
133 negative effect on insect intraspecific genetic diversity (5', 2°, 4° scales), but with no significant
134 effect on genetic diversity in other taxonomic classes (Fig. 4; Table S1). A significant positive
135 effect of the interaction between longitude and human density was detected on the genetic
136 diversity in fish (1° scale) only, while a negative effect of longitude and land use intensity was
137 identified exclusively for bird genetic diversity (4° scale). Lastly, the interaction between latitude
138 and longitude was found to have a positive effect on insect genetic diversity at the 5' spatial
139 scale (Fig. 4; Table S1). Thus, apart from latitudinal gradients and positive temporal trends in
140 bird and fish intraspecific genetic diversity, the effects of human density, land use intensity, and
141 sampling year were largely taxon- and scale-dependent.
142
143 *Times series of intraspecific genetic diversity*
144 We also examined temporal trends in individual populations that were sampled
145 repeatedly over time. However, the limited availability of populations with multiple (≥ 3) years
146 of sequence data restricted our time series analysis to the largest spatial scale of analysis (4° grid
147 cells; Table 1). Our time series dataset included 1,100 populations from 965 species and 98 grid
148 cells (mean ± se duration of time series = 3.4 ± 0.05 years, range = 3-15 years). No significant
149 mean temporal trend was detected for any taxa (Fig. 5, Table S2). This finding held when we
150 excluded very short time series (< 5 years of data) from the analysis (Fig. S2). Nonetheless,
151 diversity was apparently changing in many populations (Fig. 5), although time series were too
152 short to assess the statistical significance of these trends[26]. Models including latitude, land use
153 intensity or human population density indicated that none of these variables had a significant
154 impact on temporal trends in genetic diversity at the 4° scale (Table S2). However, we found
155 significant main effects of latitude on genetic diversity in insects and mammals (Table S2),

confirming the importance of latitude as a predictor of mean diversity, if not temporal trends in diversity. In conclusion, this time series analysis also uncovered a lot of temporal variation in genetic diversity, but it suggested that no overall trend exists at the global scale or across large anthropogenic gradients.

Discussion

Analyzing global patterns of intraspecific genetic diversity in birds, fishes, insects, and mammals, we have shown that human impacts on genetic diversity are scale are taxon dependent. We also find that the elevated intraspecific genetic diversity documented for mammals at low latitudes[25] extends to fishes, insects, and the adapted dataset of mammals in our analysis (i.e. the subset of mammalian sequences with known collection years). Over time, we observed significant increases in intraspecific genetic diversity in fishes at all spatial scales, and in birds at small spatial scales. In the time series analysis, these trends were not visible, and the average temporal trend across populations was zero for all taxa. However, some strong temporal trends at the population level were found in all taxa, indicating that intraspecific genetic diversity is a dynamic dimension of diversity which warrants further attention.

Current estimates indicate that up to 70% of the Earth's surface has been modified by human activities, largely within the last century[27]. Human impacts on the environment such as urbanization and land use intensification are known to influence intraspecific variation and species evolutionary parameters[16,28–30]. However, we did not detect significant declines in genetic diversity in areas affected (and settled) by humans, nor did we observe systematic temporal declines. The presence of humans can have both negative and positive effects on species diversity[22,24] and we also find both negative and positive effects on intraspecific genetic diversity (see also[16]). Human effects on population selection and drift are highly heterogeneous and should not be expected to generate an overall pattern of declining genetic diversity across all taxa and sites. Furthermore, the changes in magnitude and sometimes direction of human impacts across scales in our analysis confirms that the scale at which diversity is estimated can influence the conclusions we draw regarding the overall effects of human activity on genetic variation, reinforcing the necessity of scale-explicit analyses.

While our results contradict the recent finding that intraspecific genetic diversity is lower in human-dominated areas[25], we believe that detrimental effects of human activity and steady declines in global genetic diversity is an important potential outcome that is difficult to detect

188 because of several important data biases. We acknowledge that the COI locus does not evolve
189 under neutrality[31] and that sequence variation may not reflect anthropogenic pressures[32]. More
190 appropriate genetic tools and metrics are known for measuring neutral intraspecific diversity
191 (e.g., microsatellites, allelic diversity); however, there is currently no global database for these
192 data. COI is thus one of few genes with abundant sequences in common databases[33], with
193 metadata readily available (e.g. spatial coordinates and year of sequence collection), and with
194 sequences available for a large number of species (likely due to the use of COI for species
195 identification). Despite this wealth of COI data with collection years, we still noted some
196 taxonomic gaps in the database, e.g. a small number of Coleoptera sequences despite this order
197 accounting for a large number of insect (and animal) species. Another important advance for
198 future assessment of global genetic diversity trends would be to incorporate species range sizes.
199 Mismatch in the scale at which species move throughout the landscape and the scale at which
200 they are assessed can cause important relationships to go undetected[34–36].
201 Finally, more time series of genetic diversity within individual populations are urgently
202 needed. Despite the size of our sequence dataset, reliable time series analysis for all taxa could
203 only be conducted at the 4° scale. At this scale, individuals from distant sites would have been
204 grouped into a single population and this can mask temporal trends if different subpopulations
205 are genetically divergent. Even at the 4° scale, most time series were very short (< 5 years). The
206 lack of replication in time is a persistent problem in modern ecology[37] and evolutionary biology
207 that is constraining our ability to make strong inferences about global patterns of biodiversity
208 change[26,38]. We urge data collectors to upload metadata such as collection year and spatial
209 coordinates when depositing sequences in databases–a remarkably large number of sequences in
210 GenBank do not have a collection year (e.g. 95% of amphibian sequences), which constrained
211 our analyses for some taxa.
212
213 **Conclusion**
214 Anthropogenic activity has complex, scale and taxon-specific effects on intraspecific
215 genetic diversity. There is a clear opportunity to establish a global and systematic monitoring
216 program for intraspecific diversity[7]. Global monitoring of genetic diversity would improve our
217 ability to detect change and attribute the causes of worldwide patterns of spatial and temporal
218 variation in genetic diversity we report here.
219

Methods

R version 3.5.0[39] was used for all analyses described below. We combined two large datasets: one of global land use and one of animal genetic sequences, both including spatial and temporal information. Spatial coordinates were used to overlay genetic diversity and land use data on a gridded world map. To calculate population genetic diversity, we defined 'populations' as unique species × grid cell combinations (e.g. white-tailed deer in grid cell 52-56°N and 114-118°W). Our analysis excluded grid cells composed entirely of water; however, we included data from aquatic animals found in grid cells with some land in them. We reasoned that land use can have impact on inland and coastal waters, and thus on animals found in these environments[40,41]. To evaluate the scale-dependence of genetic diversity trends and human impacts, we calculated land use and genetic diversity at four spatial scales, namely 5' (0.08°), 1°, 2°, and 4° square grid cells.

Sequence data

Mitochondrial cytochrome c oxidase subunit I (mtDNA COI) sequences for birds (Aves), fishes (Actinopterygii), insects (Insecta), and mammals (Mammalia) were downloaded from the National Center for Biotechnology Information (NCBI) 'GenBank'[42] and from the 'Barcode of Life Data Systems'[43] (BOLD) in April 2017. Sequences from GenBank were retrieved with the Entrez Utilities Unix Command Line, while for BOLD we used the application platform interface. Only sequences with documented geographic coordinates and sampling dates available in the databases were downloaded. Sequences with ambiguous taxonomic assignment (e.g., species name containing '.spp') were excluded from the analysis. Species, grid cell, and year-specific sequence alignments were then performed using default parameters in MAFFT[44]. Pairwise nucleotide differences were calculated for all pairs of sequences with > 50% sequence overlap as in[25]. We then calculated the mean pairwise dissimilarity among all sequences/individuals from a population to estimate nucleotide diversity ($\hat{\pi}$), a measure of population genetic diversity. Species present in multiple grid cells in the same year were treated as independent populations, and separate $\hat{\pi}$ values were estimated accordingly. For populations with multiple years of data, separate $\hat{\pi}$ values were computed for each year. The few diversity estimates pre-1980 were discarded, as were extreme values 10 standard deviations greater than the mean of all estimates. We also gathered data for plants (markers ITS, MatK, RbcL),

amphibians (COI), and molluscs (COI), but the proportion of sequences with collection dates was very low, precluding an analysis of temporal trends.

Nucleotide diversity estimates ($\hat{\pi}$; henceforth, 'genetic diversity') were re-calculated at each spatial resolution/grid cell size. Since at least two sequences from any species-grid cell combination is required to calculate genetic diversity, it should be noted that the number of sequences in the dataset increases with scale. Aggregating sequences into lower resolutions, or relatively larger grid cells, increased the number of possible pairwise sequence comparisons, and thus population sizes and their spatial extents. In parallel, the number of populations available for each taxonomic class often diminished. Genetic diversity would be expected to increase with grid cell size as distant (potentially isolated) individuals are grouped into large populations.

Land use and human population data

Global human population and land use estimates were obtained from the most extensive and up-to-date version of the 'History Database of the Global Environment'[45] (HYDE 3.2) for years 1980-2016. HYDE 3.2 provides land use and human population density estimates for all land masses at 5' spatial resolution, with data available for every decade from 1980 to 2000, and every year after 2000. Variables included in our analyses were: maximum land area (km^2/grid cell), human population counts (inhabitants/grid cell), as well as four classes of intensive land use: cropland, pasture, converted rangeland, and built-up area (km^2/grid cell). See[45] for a detailed description of these variables.

All HYDE 3.2 global datasets for these six variables and for the time-period 1980-2016 were converted to raster data structures and aggregated at lower spatial resolutions using the R package 'raster'[46]. Values for land area, human population, and each land use category were computed for each new 1°, 2°, and 4° grid cell by summing the values of all the 5' cells of each respective variable encompassed within the new cell boundary. At each spatial resolution, human population counts were divided by the maximum land area available in each grid cell to obtain estimates of human population density (inhabitants/km^2). Similarly, cropland area, pasture area, converted rangeland area, and built-up area per grid cell were divided by land area to estimate the proportions of each cell consisting of each respective category. We then calculated an aggregate estimate of the proportion of land under intensive use for each grid cell by summing the proportions of cropland, pasture, converted rangeland, and built-up area present in each grid cell (hereafter: 'land use intensity', ranging from 0 to 1).

Large grid cells should have more heterogeneous land use, which could increase genetic diversity at the grid cell-level by creating several isolated (and potentially genetically-divergent) subpopulations. Likewise, the probability of spatial mismatch between the sequence data and the land use variables should be greater for large grid cells (e.g., the grid cell as a whole could have a large proportion of agricultural land but all animal sequences could originate from a small forest patch). We thus expected land use and human density impacts on genetic diversity to be stronger and most easily detected at small spatial scales.

Due to the absence of yearly HYDE 3.2 data for years 1980-1999, genetic diversity estimates based on sequences collected between 1980 and 1989 were assigned the 1980 human density and land use intensity values of their respective grid cell. Sequences collected between 1990 and 1999 were given 1990 values, whereas sequences sampled between 2000 and 2016 were assigned year-specific human density and land use intensity values. All data were processed using the 'tidyverse' collection of R packages[47].

Distribution of sequences and genetic diversity

We first examined the distribution of sequences in our dataset with respect to geography, taxonomy, time, human density, and land use intensity. To visualize the spatial distribution of sequences, global maps of the number of sequences per grid cell were generated at each spatial resolution using R packages 'latticeExtra'[48] and 'rworldmap'[49]. The number of sequences, populations, and species per year of each class (birds, fishes, insects, and mammals) was also tallied at all spatial resolutions. We determined whether our study populations were under levels of human influence representative of global distribution patterns by comparing the distribution of human density and land use intensity values associated with 1) all HYDE 3.2 grid cells, worldwide and pooled across all years, and 2) the time and place of sequence collections, for all spatial scales.

For each class and scale of analysis, the proportion of global taxa represented in our dataset was quantified. We first retrieved genus, family, and order-level classification for all species in our dataset using taxonomic information from the NCBI and the 'Integrated Taxonomic Information System' (ITIS) databases accessed through the R package 'taxize'[50]. Supplemental information regarding Actinopterygii order classification was obtained from 'Fishbase'[51], accessed through the R package 'rfishbase'[52]. For species absent from all of NCBI, ITIS, and Fishbase (e.g., some older/synonymous names), taxonomic information was retrieved

from BOLD itself. This order was chosen to ensure that the classification was as current as possible. We then obtained the total number of genera, families, and orders in each class from the 'Catalogue of life' database[53], also accessed through taxize, and compared these global estimates with the number of taxa in our database. This was only meant to provide a rough estimate of the proportion of taxa included in our dataset; exact proportions cannot be calculated due to unresolved taxonomy and discrepancies among taxonomic references. The proportion obtained for Actinopterygii orders (100%) is likely an over-estimate caused by large discrepancies in order number in different taxonomic databases; for example, some orders represented in BOLD/NCBI (and thus in our dataset) are present in neither Fishbase nor the Catalogue of Life and vice versa. Percentages for other classes/taxonomic levels should be more accurate as mismatch across databases was lower.

We then evaluated the spatial and temporal distribution of genetic diversity ($\hat{\pi}$) for all classes at all spatial resolutions. To map diversity in space, population-level intraspecific genetic diversity estimates were averaged across years and species to yield a single diversity value per grid cell. To visualize diversity in time, we averaged genetic diversity values within years but across all populations/grid cells to obtain a single time series of mean global genetic diversity for each animal class. Results are presented at the largest scale of analysis (4°) so that grid cells can be visually distinguished on a world map. We also plotted mean genetic diversity as a function of latitude (absolute values), longitude, land use intensity, and human population density (log-transformed), grouping observations into a small number of equal-sized bins to facilitate visualization.

Statistical analyses

Generalized linear mixed models (GLMMs) were used to assess the main and two-way interaction effects of year of sequence collection, geographical space (longitude and absolute values of latitude), human population density, and land use intensity on intraspecific genetic diversity. Models were constructed separately for each animal class and scale of analysis. We used Tweedie compound Poisson (weighted) GLMMs to accommodate the distribution of $\hat{\pi}$ values, which is continuous, positive, right-skewed, and with many exact zeros[54]. Models were fitted with the function 'cpglmm' in the R package 'cplm'[55]. All models included 'species' and 'grid cell' as random intercepts. Since most populations only had one year of data, we did not fit a random effect for population, and instead only retained one observation (the oldest in time) for

347 populations with multiple observations. In all models, the log number of sequences used to
348 calculate $\hat{\pi}$ values were used as weights, as nucleotide diversity is likely better estimated with an
349 increasing number of pairwise comparisons. Years with < 10 populations were excluded from the
350 analysis to avoid biasing temporal trends with poorly-estimated yearly averages.

351 We used a stepwise (backward) model selection procedure to identify significant
352 predictors of genetic diversity. Predictor variables included year of sequence collection, latitude
353 (absolute values), longitude, land use intensity, and human population density (log-transformed).
354 All variables were standardized to a mean of zero and a standard deviation of 1 prior to model
355 fitting. Multicollinearity was assessed with variance inflation factors[56], which were below 3 for
356 all variables in all models. For each taxonomic class and scale, we fitted a first model including
357 all possible main effects (5 parameters) and two-way interactions (10 parameters). This model
358 was subsequently reduced by removing non-significant parameters with 95% confidence
359 intervals overlapping 0, until all remaining terms were statistically-significant or contributed to
360 at least one significant two-way interaction. Increase in fit during stepwise model selection was
361 confirmed with decreases in AIC. Final models are illustrated in the main text, while the full
362 model selection results are provided in Supporting Information. Final models were validated
363 with plots of residuals against fitted values and predictor variables, with autocorrelation
364 functions (for temporal autocorrelation), and with variograms and maps of residuals (for spatial
365 autocorrelation).

366 In these models, only one data point per population was used; thus, the temporal
367 dimension of this analysis involves asking if 'year of sampling' has an effect on genetic
368 diversity. Assuming that a sufficiently large number of populations were sampled every year and
369 that mean global diversity can be properly estimated from this sample of animal populations,
370 then these spatially-averaged time series should reflect global trends in diversity. Our dataset
371 includes thousands of populations, but inconsistent taxonomic and/or spatial coverage across
372 years could still distort temporal trends. Therefore, as a perhaps more robust, but also more data-
373 limited approach, we also investigated trends in the subset of populations that were sampled
374 repeatedly over time (3 or more years of data). This analysis focused on diversity values
375 computed at the 4° scale, which provided the largest number of time series. We again used
376 weighted Tweedie compound Poisson GLMMs, including 'grid cell' and 'species' as random
377 intercepts, and log number of sequences as weights. We fitted four models per taxonomic class.
378 The first model, testing whether there is an overall temporal trend in genetic diversity, included a

379 fixed effect of 'year' and a population-specific random slope and intercept for 'year'. The results
380 of this model (fitted values and population slopes) are shown in the main text.
381 Then, we also asked whether latitude (absolute values), land use intensity, and human
382 population density (log-transformed) influenced temporal trends, by fitting three additional
383 models including one of these variables as an additional main (fixed) effect and as a two-way
384 interaction effect with 'year'. Separate models had to be fitted for this analysis because the three
385 variables were collinear in this subset of our full dataset. Parameter estimates from these models
386 are reported in Supporting Information. All models were validated as were the GLMMs using the
387 full dataset. Finally, because time series duration can influence detectability of temporal trends in
388 diversity[26], we also repeated the analysis using more stringent selection criteria for time series
389 inclusion (at least 4, 5, or 6 years of data), and verified that inferences remained unchanged. We
390 also extracted the random slopes (population trends) fitted in the models and related those to
391 time series duration, using linear regression to test for an association between the two variables.
392
393 **Data and code availability**
394 All data used in this manuscript are available online, as described in the Methods section.
395 Formatted datasets used in statistical analyses will be archived on an online repository upon
396 manuscript acceptance. Data manipulation and analysis code can be found online at
397 https://github.com/VFugere/GenDivProject.
398
399 **References**
400
401 1. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907 (2000).
402 2. Reed, D. H. & Frankham, R. Correlation between fitness and genetic diversity. *Conserv.*
403 *Biol.* **17**, 230–237 (2003).
404 3. Frankham, R. Genetics and extinction. *Biol. Conserv.* **126**, 131–140 (2005).
405 4. Bijlsma, R. & Loeschcke, V. Genetic erosion impedes adaptive responses to stressful
406 environments. *Evol. Appl.* **5**, 117–129 (2012).
407 5. Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N. & Vellend, M. Ecological
408 consequences of genetic diversity. *Ecol. Lett.* **11**, 609–623 (2008).

409 6. Pereira, H. M. *et al.* Essential biodiversity variables. *Science* **339**, 277–278 (2013).
410 7. Mimura, M. *et al.* Understanding and monitoring the consequences of human impacts on
411 intraspecific variation. *Evol. Appl.* **10**, 121–139 (2017).
412 8. Paz-Vinas Ivan *et al.* Systematic conservation planning for intraspecific genetic diversity.
413 *Proc. R. Soc. B Biol. Sci.* **285**, 20172746 (2018).
414 9. Palumbi, S. R. Humans as the world's greatest evolutionary force. *Science* **293**, 1786–1790
415 (2001).
416 10. Alberti, M. Eco-evolutionary dynamics in an urbanizing planet. *Trends Ecol. Evol.* **30**, 114–
417 126 (2015).
418 11. Thomas, C. D. Rapid acceleration of plant speciation during the Anthropocene. *Trends Ecol.*
419 *Evol.* **30**, 448–455 (2015).
420 12. Schlaepfer, D. R., Braschler, B., Rusterholz, H.-P. & Baur, B. Genetic effects of
421 anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-
422 analysis. *Ecosphere* **9**, e02488 (2018).
423 13. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
424 14. King, J. & Jukes, T. Non-Darwinian evolution. *Science* **164**, 788–798 (1969).
425 15. Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press,
426 1983).
427 16. DiBattista, J. D. Patterns of genetic variation in anthropogenically impacted populations.
428 *Conserv. Genet.* **9**, 141–156 (2008).
429 17. Banks, S. C. *et al.* How does ecological disturbance influence genetic diversity? *Trends*
430 *Ecol. Evol.* **28**, 670–679 (2013).
431 18. Dubrova, Y. E. *et al.* Human minisatellite mutation rate after the Chernobyl accident. *Nature*
432 **380**, 683–686 (1996).

433 19. Ellegren, H., Lindgren, G., Primmer, C. R. & Møller, A. P. Fitness loss and germline
434 mutations in barn swallows breeding in Chernobyl. *Nature* **389**, 593 (1997).
435 20. Bickham, J. W., Sandhu, S., Hebert, P. D. ., Chikhi, L. & Athwal, R. Effects of chemical
436 contaminants on genetic diversity in natural populations: implications for biomonitoring and
437 ecotoxicology. *Mutat. Res. Mutat. Res.* **463**, 33–51 (2000).
438 21. Crispo, E., Moore, J.-S., Lee-Yaw, J. A., Gray, S. M. & Haller, B. C. Broken barriers:
439 human-induced changes to gene flow and introgression in animals. *BioEssays* **33**, 508–518
440 (2011).
441 22. McGill, B. J., Dornelas, M., Gotelli, N. J. & Magurran, A. E. Fifteen forms of biodiversity
442 trend in the Anthropocene. *Trends Ecol. Evol.* **30**, 104–113 (2015).
443 23. Jarzyna, M. A. & Jetz, W. Taxonomic and functional diversity change is scale dependent.
444 *Nat. Commun.* **9**, 2565 (2018).
445 24. Pautasso, M. Scale dependence of the correlation between human population presence and
446 vertebrate and plant species richness. *Ecol. Lett.* **10**, 16–24 (2007).
447 25. Miraldo, A. *et al.* An Anthropocene map of genetic diversity. *Science* **353**, 1532–1535
448 (2016).
449 26. White, E. R. Minimum time required to detect population trends: the need for long-term
450 monitoring programs. *BioScience* (in press). doi:10.1093/biosci/biy144
451 27. Foley, J. A. *et al.* Global consequences of land use. *Science* **309**, 570–574 (2005).
452 28. Alberti, M. *et al.* Global urban signatures of phenotypic change in animal and plant
453 populations. *Proc. Natl. Acad. Sci.* **114**, 8951–8956 (2017).
454 29. Fugère, V. & Hendry, A. P. Human influences on the strength of phenotypic selection. *Proc.*
455 *Natl. Acad. Sci.* **115**, 10070–10075 (2018).

456 30. Hendry, A. P., Farrugia, T. J. & Kinnison, M. T. Human influences on rates of phenotypic
457 change in wild animal populations. *Mol. Ecol.* **17**, 20–29 (2008).
458 31. Pentinsaari, M., Salmela, H., Mutanen, M. & Roslin, T. Molecular evolution of a widely-
459 adopted taxonomic marker (COI) across the animal tree of life. *Sci. Rep.* **6**, 35275 (2016).
460 32. Bazin, E., Glémin, S. & Galtier, N. Population size does not influence mitochondrial genetic
461 diversity in animals. *Science* **312**, 570 (2006).
462 33. Porter, T. M. & Hajibabaei, M. Over 2.5 million COI sequences in GenBank and growing.
463 *PLOS ONE* **13**, e0200177 (2018).
464 34. Wiens, J. A. Spatial scaling in ecology. *Funct. Ecol.* **3**, 385–397 (1989).
465 35. Levin, S. A. The problem of pattern and scale in ecology. *Ecology* **73**, 1943–1967 (1992).
466 36. Chave, J. The problem of pattern and scale in ecology: what have we learned in 20 years?
467 *Ecol. Lett.* **16**, 4–16 (2013).
468 37. Estes, L. *et al.* The spatial and temporal domains of modern ecology. *Nat. Ecol. Evol.* **2**,
469 819–826 (2018).
470 38. Gonzalez, A. *et al.* Estimating local biodiversity change: a critique of papers claiming no net
471 loss of local diversity. *Ecology* **97**, 1949–1960 (2016).
472 39. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for
473 Statistical Computing, 2018).
474 40. Dudgeon, D. *et al.* Freshwater biodiversity: importance, threats, status and conservation
475 challenges. *Biol. Rev.* **81**, 163–182 (2006).
476 41. Stoms, D. M. *et al.* Integrated coastal reserve planning: making the land–sea connection.
477 *Front. Ecol. Environ.* **3**, 429–436 (2005).
478 42. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–D42 (2013).

43. Ratnasingham, S. & Hebert, P. D. N. bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol. Ecol. Notes* **7**, 355–364 (2007).
44. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
45. Klein Goldewijk, K., Beusen, A., Doelman, J. & Stehfest, E. Anthropogenic land use estimates for the Holocene – HYDE 3.2. *Earth Syst. Sci. Data* **9**, 927–953 (2017).
46. Hijmans, R. J. *raster: Geographic Data Analysis and Modeling.* R package version 2.5-8. (2016).
47. Wickham, H. *tidyverse: Easily Install and Load the 'Tidyverse'.* R package version 1.2.1. (2017).
48. Sarkar, D. & Andrews, F. *latticeExtra: Extra Graphical Utilities Based on Lattice.* R package version 0.6-28. (2016).
49. South, A. rworldmap: A New R package for Mapping Global Data. *R J.* **3**, 35–43 (2011).
50. Chamberlain, S. A. & Szöcs, E. taxize: taxonomic search and retrieval in R. *F1000Research* **2**, 191–191 (2013).
51. Froese, R. & Pauly, D. *FishBase.* Available at: www.fishbase.org. (2018).
52. Boettiger, C., Lang, D. T. & Wainwright, P. C. rfishbase: exploring, manipulating and visualizing FishBase data from R. *J. Fish Biol.* **81**, 2030–2039 (2012).
53. Species 2000 & ITIS Catalogue of Life, 2018 Annual checklist. Digital resource at www.catalogueoflife.org/annual-checklist/2018. (2018).
54. Dunn, P. K. & Smyth, G. K. Series evaluation of Tweedie exponential dispersion model densities. *Stat. Comput.* **15**, 267–280 (2005).
55. Zhang, Y. Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Stat. Comput.* **23**, 743–757 (2013).

56. Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R*. (Springer New York, 2009).

Acknowledgements

This work resulted from insightful discussions among past and present Gonzalez lab members. We are grateful to Andreia Miraldo and Michael Borregaard for assistance with Julia scripts. We thank Navin Ramankutty, Erle Ellis, and Kees Klein Goldewijk for help accessing and interpreting the HYDE 3.2 dataset. The authors also acknowledge support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Quebec Centre for Biodiversity Science (QCBS). AG is supported by the Liber Ero Chair in Biodiversity Conservation and a Killam Fellowship.

Author Contributions

All authors contributed to the idea and design of the project. KLM and FJJC contributed to sequence data collection and analysis. VF led statistical analyses. CD led GIS/spatial analyses. AGr collected taxonomic information. KLM drafted the manuscript. VF and CD made the figures. AGo supervised the project. All authors contributed to data interpretation and writing of the manuscript.

Competing interests

The authors declare no competing interests.

Table 1. Number of COI sequences, populations, species, genera, families, orders, and time series included in the dataset.

scale	class	sequences	populations	species	genera	families	orders	time series
5’	Aves	6337	2139	1320 (12.75%)	685 (30.67%)	120 (52.86%)	31 (77.50%)	3
5’	Actinopterygii	19679	4935	2453 (7.54%)	1074 (34.17%)	251 (51.43%)	46 (100%)	40
5’	Insecta	103102	19743	9603 (1.12%)	4371 (5.71%)	428 (37.88%)	24 (85.71%)	114
5’	Mammalia	16974	2619	560 (9.57%)	232 (18.04%)	45 (28.48%)	11 (37.93%)	25
1°	Aves	7016	2302	1451 (14.01%)	724 (32.41%)	120 (52.87%)	31 (77.50%)	6
1°	Actinopterygii	20733	4692	2538 (7.81%)	1103 (35.09%)	254 (52.05%)	46 (100%)	68
1°	Insecta	112008	18997	10566 (1.23%)	4658 (6.08%)	439 (38.85%)	24 (85.71%)	540
1°	Mammalia	17713	1988	572 (9.78%)	238 (18.51%)	46 (29.11%)	12 (41.38%)	74
2°	Aves	7245	2332	1497 (14.46%)	737 (33.00%)	120 (52.86%)	31 (77.50%)	15
2°	Actinopterygii	21100	4460	2559 (7.87%)	1106 (35.19%)	255 (52.25%)	58 (126.09%)	84
2°	Insecta	114858	18781	10940 (1.27%)	4761 (6.22%)	443 (39.20%)	24 (85.71%)	680
2°	Mammalia	17932	1665	577 (9.86%)	238 (18.51%)	46 (29.11%)	12 (41.38%)	122
4°	Aves	7577	2350	1547 (14.94%)	753 (33.71%)	122 (53.74%)	31 (77.50%)	36
4°	Actinopterygii	21340	4176	2577 (7.93%)	1110 (35.32%)	256 (52.46%)	46 (100%)	112
4°	Insecta	117474	18273	11240 (1.31%)	4844 (6.33%)	445 (39.38%)	24 (85.71%)	820
4°	Mammalia	18127	1397	582 (9.95%)	239 (18.58%)	46 (29.11%)	12 (41.38%)	135

Populations are defined as unique species × grid cell combinations while time series represent populations with 3 or more years of data. Scale indicates grid cell dimension (side length). Percentages in parenthesis for species, genera, families and orders are estimates of the proportion of taxa present in the database relative of the total number of taxa globally, as described in the Methods. Note that the percentage for Actinopterygii orders is an over-estimate caused by large discrepancies in order number in different taxonomic databases.

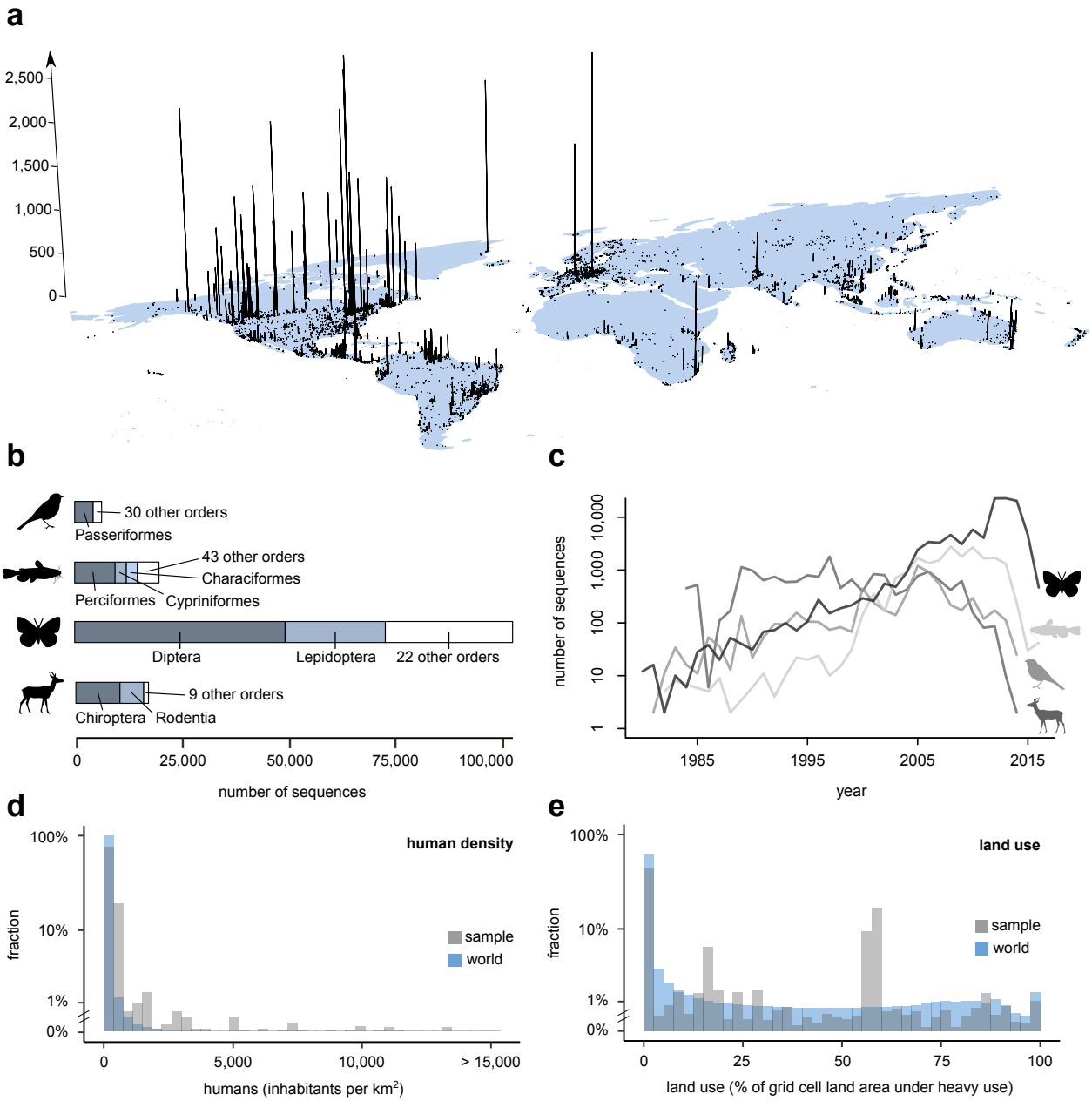

Figure 1. Distribution of COI sequences used to compute genetic diversity at the 5' spatial resolution across geographical space (a), taxonomic classes (b), time (c), and anthropogenic parameters (d, e). (a) Global distribution of sequences available in NCBI or BOLD databases with known geographic coordinates and year of sequence collection. Bar heights represent the total number of sequences per 5' grid cell (min=2 sequences, max=2,947) from the four animal classes. (b) Distribution of sequences across the four taxonomic classes: birds (N=6,337), fish (N=19,679), insects (N=103,102), and mammals (N=16,974 sequences). Orders contributing a large proportion of sequences are indicated. (c) Number of sequences available in the dataset for each year and animal class. (d,e) Distribution of sequences (grey) according to human density (d) and land use intensity (e) relative to the frequency of these parameter values worldwide (blue).

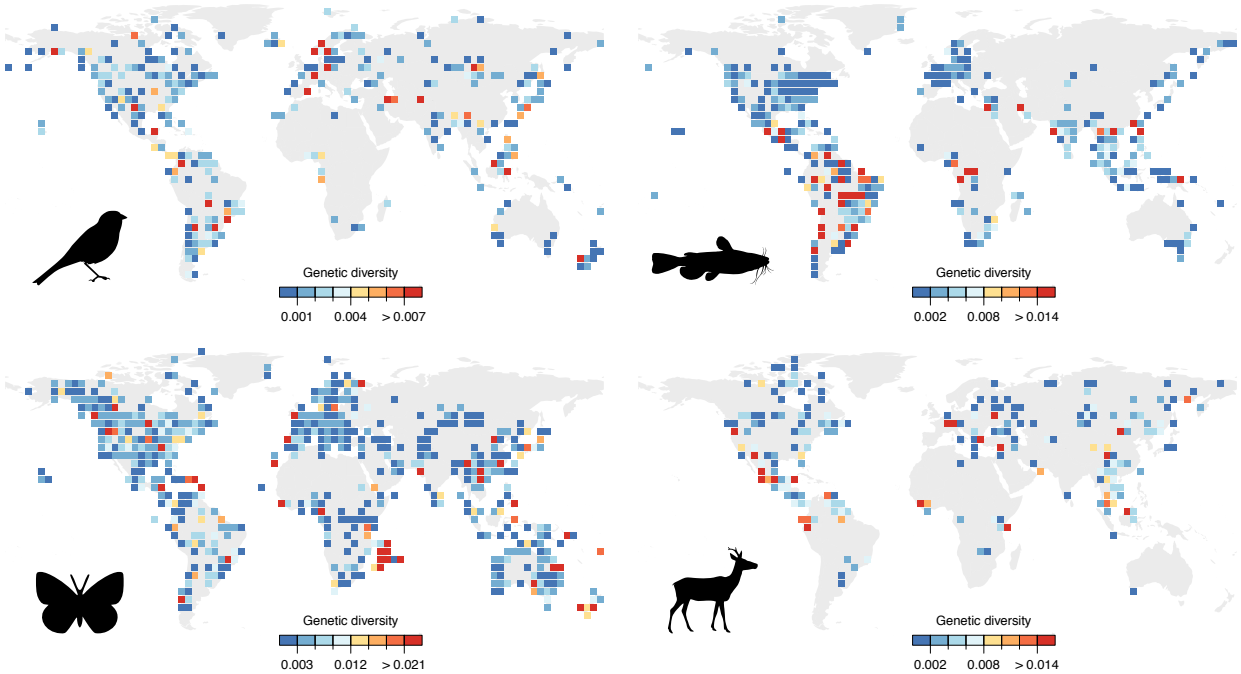

Figure 2. Spatial variation in intraspecific genetic diversity of birds, inland and coastal bony fishes, insects, and mammals. Genetic diversity estimates were averaged across populations and years to yield a single diversity value per grid cell (shown here at the 4° resolution). Note that genetic diversity scales differ among classes.

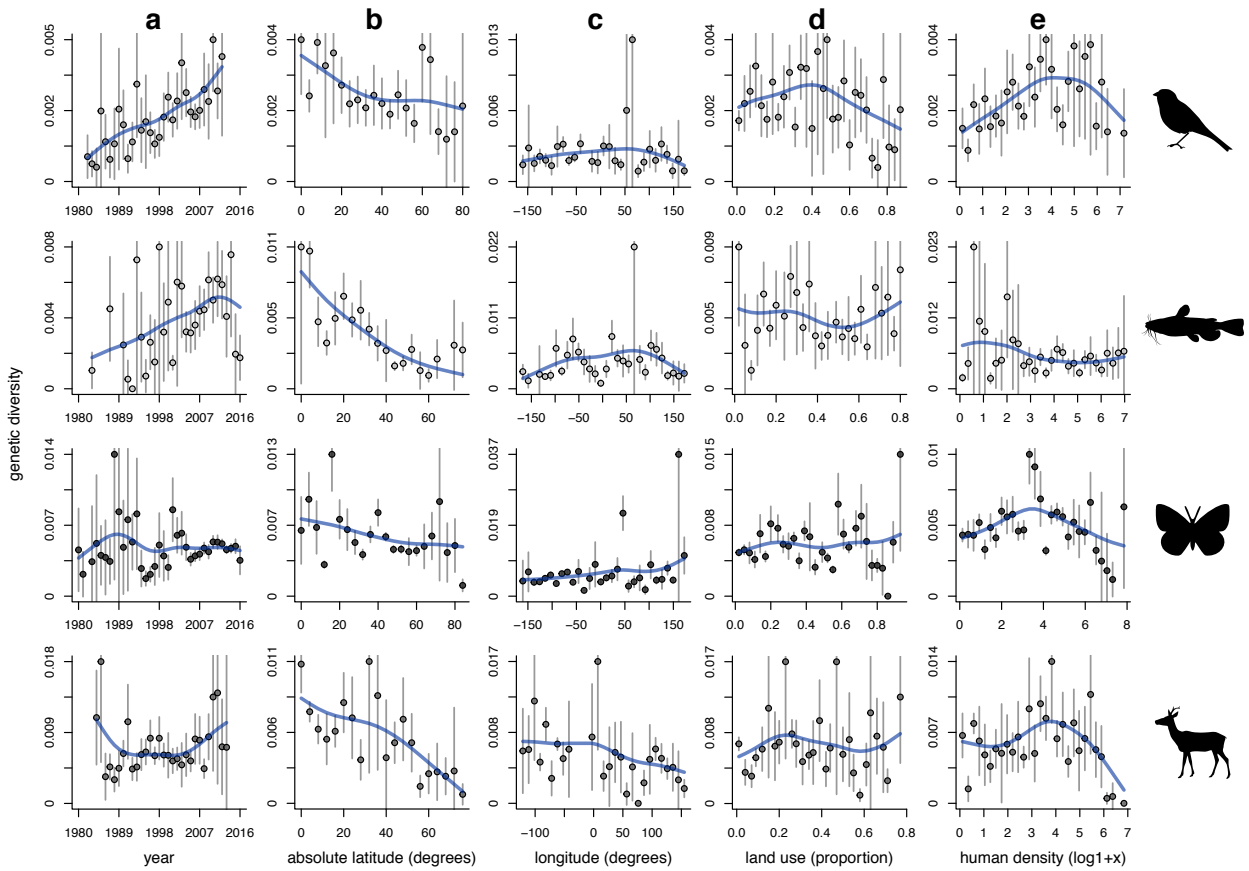

Figure 3. Genetic diversity calculated in 4° grid cells as a function of five of its potential drivers. (a) Mean intraspecific genetic diversity of each class over time. Population-level genetic diversity values from any given year were averaged across all grid cells to yield spatially-averaged global time series. Different populations/grid cells contribute to each yearly average. (b-e) Relationship between genetic diversity and grid cell latitude (b), longitude (c), land use intensity (d), and human population density (e). Estimates from all years were pooled, and data were binned into a small number of groups for illustration purposes. In all panels, symbols and error bars indicate mean ± 95% confidence intervals, while thick lines are cubic splines weighted by the log number of diversity estimates included in each data point. To aid visualization, values of confidence intervals lower than 0 (the minimum theoretical value for nucleotide diversity) or larger than the maximum mean value are not shown.

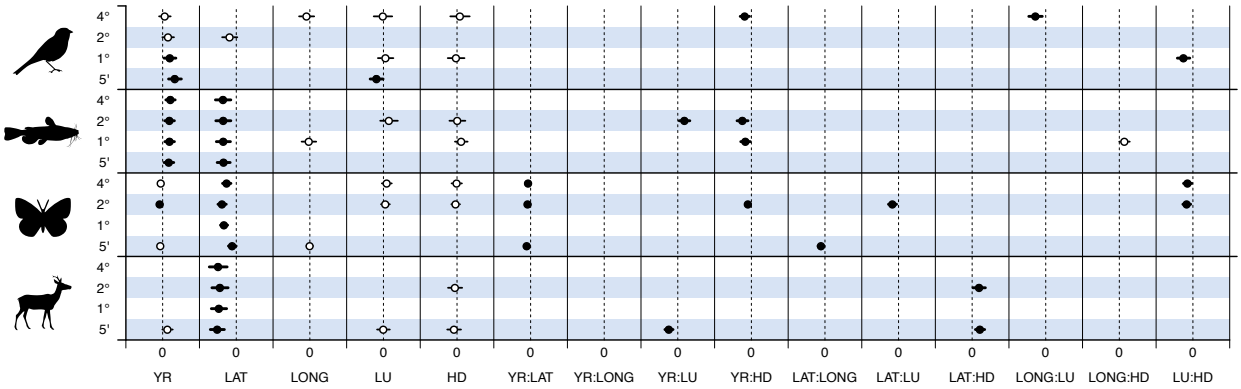

Figure 4. Effects of year of sampling, latitude (absolute values), longitude, land use intensity, and human population density (log-transformed) on the genetic diversity of populations of birds, inland and coastal bony fishes, insects, and mammal. Symbols indicate main effects and two-way interaction effects estimated by GLMMs fitted independently for each taxon and scale; negative and positive values respectively decrease and increase genetic diversity. Error bars indicate 95% confidence intervals of parameter values. Only effects retained in final models after stepwise model selection are shown (see Table S1 for all model selection results). Filled symbols represent significant effects with confidence intervals that do not overlap zero, while open symbols indicate non-significant (main) effects retained in final models because they contribute to at least one significant two-way interaction. YR: year. LAT: latitude. LONG: longitude. LU: land use intensity. HD: human population density.

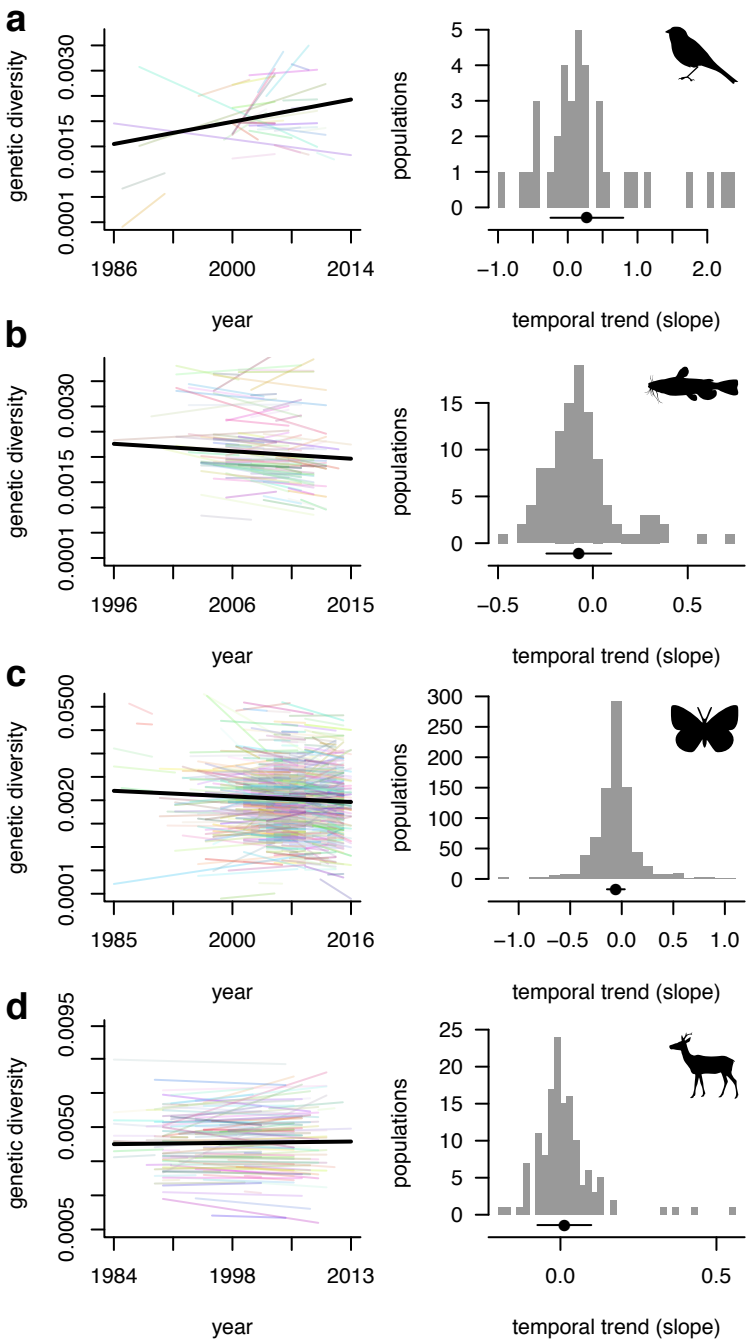

Figure 5. Time series analysis of genetic diversity in individual populations of birds (a), inland and coastal bony fishes (b), insects (c), and mammals (d) in 4° grid cells. Left panels illustrate random slopes models in which each population has a different intercept and slope for the effect of time on diversity. Thick black lines indicate overall trends across populations (see Table S2), while thin colored lines show fitted values for individual populations (slopes of these lines represent 'population trends'). Right panels show the distribution of estimated population trends. Symbols below histograms indicate overall (fixed) effects of year on genetic diversity across time series (error bars = 95% confidence intervals for parameter estimate).