

## Supplementary Materials for

### An Anthropocene map of genetic diversity

Andreia Miraldo,\* Sen Li, Michael K. Borregaard, Alexander Flórez-Rodríguez,  
Shyam Gopalakrishnan, Mirnesa Rizvanovic, Zhiheng Wang, Carsten Rahbek,  
Katharine A. Marske, David Nogués-Bravo\*

\*Corresponding author. Email: [andreia.miraldo@snm.ku.dk](mailto:andreia.miraldo@snm.ku.dk) (A.M.); [dnogues@snm.ku.dk](mailto:dnogues@snm.ku.dk) (D.N.-B.)

Published 30 September 2016, *Science* **353**, 1532 (2016)

DOI: [10.1126/science.aaf4381](https://doi.org/10.1126/science.aaf4381)

#### This PDF file includes:

Materials and Methods

Figs. S1 to S19

Tables S1 to S5

References

#### Other supplementary material for this manuscript includes the following:

Data file S1: R code and data

Data file S2: Functionality of web tool iMapGenes

## **Materials and Methods**

### **Retrieving sequence data from GenBank and BOLD**

We created our mitochondrial DNA database by downloading all sequence records and all associated annotations (locality name, geographic coordinates, loci identification and taxonomic identification) for both mammals and amphibians available in GenBank ([www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank)) and BOLD ([www.boldsystems.org](http://www.boldsystems.org)). GenBank data were downloaded on the 14th of December 2014 using the Entrez Utilities Unix Command Line. Data from BOLD were downloaded on the 7th of September 2015 directly from BOLD webpage using the application-platform interface (API) provided ([http://www.boldsystems.org/index.php/API\\_Public/](http://www.boldsystems.org/index.php/API_Public/) combined?). This resulted in a total of 318,748 mitochondrial sequences for 7,125 species: 236,056 sequences for 3,930 species of mammals and 82,692 sequences for 3,195 species of amphibians (Table S1) - hereafter the “full database”. The 318,748 records in our full database consist of sequences of different lengths and loci, from partial gene sequences with less than 100 base pairs (bp) to complete mitochondrial genomes (>15,000 bp).

### **Georeferencing sequences using locality annotations**

Of the 318,748 sequences downloaded, 40,703 (13%) were already annotated with geographic coordinates (latitude and longitude). To assign geographic coordinates to the remaining sequences we used the API tool provided by GeoNames.org (<http://api.geonames.org>), which automatically assigns latitude and longitude to locality names. To increase the quality of the data in our georeferenced database, we excluded from this procedure any sequences that had only the country name as locality information. We were able to assign geographical coordinates to 60,088 sequences (65% success rate) out of the 92,449 sequences with some locality information provided. Our georeferenced sequence database is thus composed of 100,791 records, representing approximately 32% of the sequences obtained from GenBank and BOLD (Table S1).

### **Quality control of georeferenced sequences**

To measure the accuracy of the assignment of geographic coordinates from locality data using GeoNames.org, we used the 13% of sequences from the full database that were retrieved from GenBank/BOLD already annotated with both locality names and geographical coordinates (21,825 sequences). For each sequence we then obtained new geographic coordinates using the locality annotations following the procedure described above. To compare the accuracy of these newly assigned geographic coordinates, we calculated for each sequence the distance between the original geographic coordinates and those assigned with GeoNames.org. We transformed the distance to kilometers using the haversine formula. Results show that 90% of the localities assigned using GeoNames.org were less than 441 km from the localities originally uploaded as annotations, with an average distance of 216 km (s.d. 926 km). This information was used to choose the best grid cell area for downstream analyses of spatial data to minimize the effects of georeferencing errors.

For each species we also assessed the spatial concordance between the georeferenced sequences and species' geographic ranges. We used the gridded IUCN species ranges to generate a presence/absence matrix with names of species as columns and equal-area grid cells as rows (see "Gridding IUCN species ranges" section below). For each species, we then assessed whether the georeferenced sequences fell within or outside species ranges. Our results showed that 90% of georeferenced sequences fell within species ranges and another 3% fell in grid cells immediately adjacent to species ranges. Overall, our quality control procedures indicated that the method used to assign geographic coordinates to the 60,088 previously unannotated sequences is powerful and accurate.

### **Retrieving single locus sequences**

Because mutation rates vary considerably among loci, we opted to estimate global genetic diversity (see below) using a single-locus approach. We chose *Cytochrome b* (*cytb*) as it was the locus that maximized the number of georeferenced sequence data available in our database across taxa (Table S2). From the full set of georeferenced sequences, we first extracted all sequences that were annotated as *cytb*, using customized Shell scripts. In order to extract all base pairs corresponding exclusively to *cytb*, we then mapped all *cytb*-annotated sequences against a reference mitochondrial genome: *Lepus europaeus* (accession number: AJ421471) for mammals and *Typhlonectes natans* (accession number: NC002471) for amphibians. Sequence mapping was performed in Geneious v8.1.7 (31) using default settings. We then extracted all base pairs that successfully mapped against the homologous *cytb* locus of the reference genomes. The majority of sequences (>95%) mapped successfully to the reference genomes, with the exception of larger sequences (usually sequences that included more than one locus and were larger in size than the total length of *cytb* gene). For each unsuccessfully mapped sequence we extracted all *cytb* base pairs using customized Shell scripts. Manually extracted sequences were added to the final *cytb* database. To assess the sensitivity of our genetic diversity results to the choice of locus, we also extracted all sequences of *cytochrome oxidase subunit 1* (*co1*) gene for mammals. We chose *co1* because it is the locus with the highest number of georeferenced sequences for mammals in our full database (Table S2).

### **Sequence data preparation**

A strict filtering step was conducted to increase the quality of our database. First, we removed all sequences with IUPAC ambiguity codes as these might represent either sequencing errors or mitochondrial sequences that have been transferred to the nuclear genome (Numts, (32)). Next, we excluded sequences from domesticated, managed, laboratory reared and extinct species (see Table S3 for a complete list of excluded taxa). Following these filtering steps, species-specific alignments were generated using default settings in MUSCLE (33) and alignments were manually checked for errors using Geneious v8.1.7 (31). The majority of alignment errors detected were due to the presence of at least one non-overlapping sequence. In these specific cases, sequences were re-aligned to a reference sequence of the correspondent gene (*cytb* or *co1*) from the correspondent reference genome (see above) using the Geneious alignment algorithm (31). After re-alignment, the *cytb* (or *co1*) reference sequence was removed before downstream analyses. Finally, all sequences without georeferences were discarded before

calculating genetic diversity for each species and locus at various spatial resolutions: (1) equal-area grid cells with grain size of 385.9 km, (2) 10° latitudinal bands, (3) anthromes and (4) biomes (see below).

### **Analysis source code**

The source code used for all analyses described from this point onward is available as a ZIP archive (Additional file S1). The archive contains source files written for *julia* (34) and *R* (35), and includes a README file detailing how to create all figures and analyses presented in the main text and supplementary information.

### **Calculating Genetic Diversity (GD)**

In population genetic studies that deal with mitochondrial sequence data, the most widely used statistics to describe genetic diversity are i) haplotype richness (HR), representing the total number of unique haplotypes in a population; ii) haplotype diversity ( $h$ , (36)), a metric that incorporates information about the number of haplotypes and their relative frequency at the population level and iii) nucleotide diversity ( $\Pi$ , (37, 38)) the average number of nucleotide differences per site between two randomly chosen sequences from the same population. While each of these characterizes genetic diversity in a slightly different way, all assume that the sequences included are a representative sample of the population of origin; i.e., haplotype frequency within the sample reflects that of the parent population. Due to variation in how data are uploaded to GenBank, with accessions either representing unique specimens, haplotypes per unique sampling locality, or unique haplotypes irrespective of sampling locality, we are unable to recreate haplotype frequencies based on available information. Further, the majority of sequences in species-specific alignments are of different lengths, making the identification of haplotypes unreliable. Because of these limitations, we cannot calculate HR or  $h$ . We therefore implement nucleotide diversity as our genetic diversity metric, but with a slight modification to account for different coverage at each nucleotide site due to different sequence lengths in species-specific alignments. Nucleotide diversity is calculated as the average number of variable sites in each pairwise sequence comparison. It is mathematically defined by:

$$\Pi = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_{ij}$$
Equation 1

where  $k_{ij}$  is the number of different nucleotides between sequence  $i$  and sequence  $j$ , and  $\binom{n}{2}$  represents the number of pairwise comparisons made. Because  $\Pi$  depends on the length of the sequence ( $m$ ),  $\Pi$  is divided by  $m$  to obtain nucleotide diversity per site (38). However, in alignments composed of sequences with different lengths, dividing  $\Pi$  by  $m$  would give incorrect values of nucleotide diversity per site because each pairwise sequence comparison has a different length. To account for this, we divide  $k_{ij}$  above by the length of the sequences in each pairwise comparison instead. Not all sequences in each alignment overlap completely and therefore only pairwise comparisons where

sequences overlap in at least 50% of the longer sequence were considered. By setting this sequence overlap threshold we avoid giving too much weight to pairwise comparisons that are based on very few base pairs, which, if included, could bias nucleotide diversity values. All positions with gaps were ignored in the calculations of nucleotide diversity per site. This gives us:

$$\hat{\Pi} = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{k_{ij}}{m_{ij}}$$
Equation 2

where  $m_{ij}$  is the number of shared base pairs between sequence  $i$  and  $j$ . For a particular assemblage of species, genetic diversity was calculated by averaging nucleotide diversity per site across all species present in that assemblage. Thus, we calculate the genetic diversity of the assemblage  $t$  ( $GD_t$ ) with genetic sequences from  $S$  species as:

$$GD_t = \frac{1}{S} \sum_{p=1}^S \hat{\Pi}_{pt}$$
Equation 3

Genetic diversity was calculated at four spatial resolutions:

1) Equal area grid cells with grain size of 385.9 km. To map the distribution of genetic diversity across the planet, we projected our georeferenced aligned sequence into an equal area projection (Behrmann) and overlaid them with an equal area grid with grain size of 385.9 km representing 148,953 km<sup>2</sup> area grids (see section “Choosing grid cell grain size” for a justification of this choice). We assigned each georeferenced sequence to a grid cell using the *join attributes by location* function in QGIS (39). For each grid cell, we then calculated nucleotide diversity per site (Equation 2) independently for each species present in that grid cell as described above. The GD of each grid cell was obtained by averaging  $\hat{\Pi}$  across all species present (Equation 3). GD values for *cytb* in each grid cell were calculated independently for mammals (Fig. 1B), amphibians (Fig. 1C), and both together (Fig. 1A). To assess the sensitivity of our results to the choice of genetic marker we conducted the same analyses for mammals using another locus (*co1*). Although the spatial distribution of genetic diversity for *co1* in mammals (fig. S1) is not congruent with the spatial patterns observed for *cytb* gene at individual grid cells (fig. S1B,  $r = 0.094$ ,  $P = 0.151$ ), patterns are highly congruent for latitudinal bands (see below).

2) 10° latitude bands. To map the latitudinal distribution of genetic diversity, we assigned each georeferenced aligned sequence to a 10° latitudinal band. GD for *cytb* in each latitude band were calculated independently for mammals (Fig. 3A) and amphibians (Fig. 3C). We also calculated GD for *co1* in mammals in each latitude band (fig. S4A). Results for *co1* are, again, highly congruent with the patterns described using *cytb* (fig. S4A and Fig. 3A,  $r = 0.79$ ,  $P = 0.001$ ).

3) Anthromes. The extent of anthropogenic transformation of global land surfaces has been quantified based on human population densities and land uses, leading to a classification of the world into anthropogenic ecosystems, or anthromes (29). We use here the broadest classification of anthromes, where the degree of human transformation varies across 6 major types, from dense settlements, villages, rangelands, semi-natural lands to wildlands. Data for the anthropogenic biomes were downloaded from (<http://ecotope.org/anthromes/v2/data/>) as ASCII grids (29). The anthrome layers were provided in 5 arc minute grid. This grid was resampled at a  $1^\circ \times 1^\circ$  resolution using Block Statistics, applying a majority rule, and then polygonised (ESRI Spatial Analyst/Conversion Tools). We then classified each georeferenced aligned sequence according to the 6 anthromes using *join attributes by location* function in QGIS (39). GD values for *cytb* in each anthrome were calculated independently for mammals (Fig. 3B) and amphibians (Fig. 3D). We also calculated GD values for *co1* in mammals for each anthrome (fig. S4B).

4) Biomes. Data for the terrestrial ecoregions of the world were downloaded from World Wildlife Fund (<http://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world>) and consisted of 867 ecoregions in 14 biomes (40), provided as shapefile polygons. We classified each georeferenced aligned sequence according to one of the 14 biomes using *join attributes by location* function in QGIS (39). GD values for *cytb* in each biome were calculated independently for mammals (fig. S2) and amphibians (fig. S3), following the same procedure as described above for the equal area grid cells. We also calculated GD values for *co1* in mammals for each biome (fig. S4C). Results for both *cytb* and *co1* in mammals are highly congruent (fig. S2 and S4C).

We generated confidence intervals for the GD values of the latitudinal bands, anthromes and biomes using a bootstrapping procedure. These confidence intervals express the uncertainty in GD deriving from inter-specific differences in  $\hat{\Pi}$ . To generate confidence intervals, we resampled the  $\hat{\Pi}$  values from each species 100 times with replacement and calculated the standard deviation of the resulting GD. These are shown as error bars in the plots in Fig. 3 of the main text and figs. S2-S4.

#### **Assessing the sensitivity of GD to spatial bias in sampling intensity**

The intensity of sampling varies substantially among different grid cells in the dataset, both in terms of the number of species sampled and the number of genetic sequences sampled from each species. To evaluate the impact of the sample size of sequences on our estimates of GD, we performed several sensitivity analyses using partial resampling of the empirical data.

#### *The impact of number of sequences on estimated GD.*

To assess the behaviour of the  $\hat{\Pi}$  metric with increased sampling, we performed a rarefaction analysis using the data for the two most intensively sampled species: *Myodes glareolus*, which has the highest number of georeferenced localities in our database, and *Rana sauteri*, which has the largest number of sequences per grid cell. In each species, for each value of  $n$  between 2 and the total number of sequences, we randomly extracted

$n$  sequences and calculated  $\hat{\Pi}$  from this subset. This procedure was repeated 100 times. The results demonstrate that the expected value of  $\hat{\Pi}$  is stationary, and thus GD values are not systematically affected by sample size. However, the variance of the distribution is large at small sample sizes (fig. S5). This means that estimates of  $\hat{\Pi}$  deriving from species with very few sequences in a grid cell are associated with high uncertainty, opening the possibility that unusually high or low GD in such grid cells may be over- or underestimates deriving from insufficient sampling.

#### *Spatial reanalysis*

To evaluate the potential impact of uneven sampling of genetic sequences on our reported patterns of GD, we re-analyzed the data presented in the main text by sub-setting the data and comparing the new results to those presented in the main text via correlation analyses corrected for spatial autocorrelation (see “Statistical analysis” section below). All analyzes were done at the 385.9 km grain size grid cell level.

Exclusion of species with low sample sizes: The spatial distribution of species with low sample sizes is shown in fig. S6, which indicates, for each grid cell, the number of species that have fewer than 10 sequences (fig. S6A) and fewer than 5 sequences (fig. S6B). The spatial distribution of species with low sample sizes is global, and thus does not lead to spatial bias in the results (fig. S6A and S6B). To evaluate the impact of species with few sequences per grid cell on GD, we identified all species with fewer than 10 sequences per grid cell, and removed those species from the calculation of GD. The genetic diversity map based on this reduced dataset (fig. S7A) had a high degree of correspondence with Fig. 1A ( $r = 0.67, P < 0.001$ ). We repeated this step for all species with fewer than 5 sequences per grid cell, and the resulting genetic diversity map (fig. S8A) also had a high degree of correspondence with Fig. 1A ( $r = 0.720, P < 0.001$ ), suggesting that the overall pattern of GD is not drastically changed when we exclude species with low sample sizes from our analyses.

Rarefaction: To investigate the GD pattern expected under spatially uniform sampling, we conducted a rarefaction analysis in which we subsampled the number of sequences per species per grid cell to exactly 10. The algorithm has two steps: (1) we randomly selected exactly 10 sequences for each species with  $>10$  sequences in a given grid cell; (2) we calculated GD from this random subsample. We repeated steps (1) and (2) 100 times, and calculated the mean (fig. S7B) and standard deviation (fig. S7C) of the 100 GD values in each grid cell. We then repeated the rarefaction analysis using 5 sequences per grid cell (mean GD and standard deviation from rarefaction analysis shown in fig. S8B and S8C, respectively). The results show that the mean GD values of the 100 repetitions were very consistent with the observed GD values for species with  $>10$  sequences in each grid cell (fig. S7D,  $r = 0.999, P < 0.001$ ) and for species with  $>5$  sequences in each grid cell (fig S8D;  $r = 0.999, P < 0.001$ ). Moreover, the results did not deviate qualitatively from the analysis presented in Fig. 1A, though by necessity many fewer localities were included in the analysis. Together, these results suggest that sample size has little influence on the estimation of GD, and that the overall pattern of genetic diversity is robust to sampling differences among species.

### *An alternative metric based on sample size-weighted means*

An alternative approach to ameliorate the sensitivity of our results to sample size is to weight the contribution of each species to grid cell GD by sampling intensity. This approach uses the number of sequence pairs used to calculate  $\widehat{\Pi}$  for that species to weight the GD of the grid cell

$$\widehat{GD}_t = \frac{1}{N} \sum_{p=1}^S \widehat{\Pi}_{pt} * \binom{n_p}{2}$$

Equation 4

where  $N$  is the total number of pairwise comparisons made in the grid cell, and  $n_p$  is the total number of sequence pairs for species  $p$  in that grid cell.

This approach yielded GD values that are again highly concordant with the genetic diversity values presented in Fig. 1A (fig. S9A and S9B;  $r = 0.892$ ,  $P < 0.001$ ).

We repeated the analyses above but this time weighting the contribution of each species to grid cell GD using the total number of base pairs used to calculate  $\widehat{\Pi}$  for that species to weight the GD of the grid cell

$$\widehat{GD}_t = \frac{1}{N} \sum_{p=1}^S \widehat{\Pi}_{pt} * \binom{n_s}{2}$$

Equation 5

where  $N$  represents in this case the total number of base pairs used in the grid cell, and  $n_p$  the total number of base pairs for species  $p$  in that grid cell.

This approach yielded GD values that are again highly concordant with the genetic diversity values presented in Fig. 1A (fig. S9C and S9D;  $r = 0.956$ ,  $P < 0.001$ ).

### **Geographic representation of sequence availability and taxonomic coverage**

To map the spatial distribution of sequence availability and taxonomic coverage, we created maps of ignorance by plotting the total number of base pairs (sequence availability) against the proportion of species for which sequences are available (taxonomic coverage) per grid cell. Ignorance maps were created for all mitochondrial loci together (Fig. 2) and for *cytb* alone (fig. S10). Taxonomic coverage was calculated as the percentage of species for which we have georeferenced sequence data, relative to the species richness in each grid cell. Species richness was obtained by overlaying species ranges downloaded from IUCN Red List of Threatened Species website (41, 42), after transforming species range data from polygons to grids (see below). To compare sequence availability against taxonomic coverage we first matched all species names from GenBank and BOLD sequences to the IUCN taxonomic nomenclature using the IUCN Global Assessment search tool (<http://www.iucnredlist.org/>). For amphibians the IUCN Red List follows the nomenclature from *Amphibian Species of the World* by Frost (43), and departs from this nomenclature only in very well justified cases. The default taxonomy for mammals on the IUCN Red List is the 3rd edition of *Mammal Species of the World — A Taxonomic and Geographic Reference* based on (44). The very few cases

where we could not match species names from GenBank or BOLD and the IUCN usually represented species recently described and awaiting IUCN assessment. For those species we kept the original names from the sequence database. In a few cases, two or more species from GenBank and BOLD databases were matched to one single species in IUCN (Table S4). In these cases, species were treated as one single species, following the taxonomic nomenclature of IUCN.

### **Gridding IUCN species ranges**

All species ranges are provided by IUCN as individual polygons in an ESRI-formatted shapefile, which we gridded into an equal area grid to estimate species richness in each grid cell. IUCN range polygons are coded by “presence” (whether the species is extinct), and “origin” (whether it represents the native distribution of the species). Range polygons with presence values 3 (“possibly extant”) and 6 (“presence uncertain”) were excluded, as these represent highly uncertain records. We also excluded range polygons with origin values 3 (“introduced”) and 4 (“vagrant”), as we were interested in the natural range of each species. All marine mammals were excluded from the mammal dataset. The polygons were then superimposed on an equal area grid of grain size 385.9 km, and each species was scored as present in a given grid cell if any part of its range overlapped that grid cell (irrespective of how large the area of overlap was). The overlay analysis was done in R-3.1 (35) using the script included as additional file S1. The resulting presence-absence matrix was then transferred to the “*nodiv*” R package (45) for processing and mapping. For species that have no distribution data at the IUCN, but for which we have georeferenced sequence data, we scored the species as present in grid cells based on the geographic coordinates of the sequence data, so that it was included in the richness total.

### **Correlation analysis**

All correlations reported in the manuscript were evaluated using Pearson’s correlation coefficients. The significance of correlations between spatial patterns were estimated using Dutilleul’s (46) spatially corrected degrees of freedom. This method modifies the effective degrees of freedom by a normalization factor estimated from the degree of spatial autocorrelation in the variables. Spatial correlations and their significance were assessed using “*SpatialPack*” R package (47).

### **Effect of species range size on GD**

To examine the relationship between species range size and the number of georeferenced sequences, we performed a Pearson correlation. Results show that there is a positive relationship between range size and the number of sequences available (fig. S11, mammals: 0.187,  $P < 0.001$ ; amphibians  $r = 0.31$ ,  $P < 0.001$ ). Large range species are likely to be sampled more intensively and they might be contributing disproportionately more towards the genetic diversity pattern presented here. This would be problematic, especially if species with larger ranges have consistently higher (or lower) values of  $\widehat{\Pi}$  than smaller range species. To check if  $\widehat{\Pi}$  differs between large and small range species we contrasted  $\widehat{\Pi}$  values from species belonging to the top 90<sup>th</sup> percentile in terms of range size to the bottom 10<sup>th</sup> percentile. Comparison between groups was done using a non-parametric test (Kruskal-Wallis). Results show that at a grid cell level  $\widehat{\Pi}$  does not

differ between groups (amphibians:  $\chi^2 = 2.9837$ ,  $P = 0.08411$  mammals:  $\chi^2 = 0.0156$ ,  $P = 0.9007$ ; fig. S12).

### Relationship between genetic diversity and latitude

We investigated the effects of latitude on genetic diversity at two levels: (1) grid cell level and (2) latitudinal band level. For individual grid cells, we performed a beta regression linear model (48), with absolute latitude as the predictor variable and genetic diversity as the response variable. In order to remove values of 0 in the response variable, we first transformed genetic diversity values using the formula  $(y.(n-1)+0.5)/n$ , where  $n$  is the sample size (49). Beta regression models were run in the R package “*betareg*” (50). We performed independent models for amphibians, mammals, and both taxa combined using the *cytb* locus, and again for mammals using the *co1* locus. In all cases, results show a significant linear relationship between genetic diversity and absolute latitude, with negative slopes, indicating that genetic diversity decreases as absolute latitude increases (fig. S13; ). For the latitudinal bands, we fit a beta regression linear model (second degree polynomial function with a quadratic term) with latitudinal band centroid as the predictor variable and genetic diversity, transformed as above, as the response variable. We ran independent models for amphibians and mammals with *cytb* and for mammals with *co1*. In all cases we found a significant hump-shaped quadratic relationship (amphibians: pseudo  $r^2 = 0.73$ ; mammals *cytb*: pseudo  $r^2 = 0.80$ ; mammals *co1*: pseudo  $r^2 = 0.89$ ) between latitudinal band and genetic diversity, with significant negative quadratic terms (amphibians:  $\beta_2 = -4.74^{e-4}$ ,  $P < 0.001$ ; mammals *cytb*:  $\beta_2 = -3.41^{e-4}$ ,  $P < 0.001$ ; mammals *co1*:  $\beta_2 = -4.06^{e-4}$ ,  $P < 0.001$ ).

### Relationship between genetic diversity and anthromes

To investigate whether genetic diversity differs among anthromes we carried out a Kruskal-Wallis test (KW-test) independently for amphibians and mammals using the *cytb* gene and mammals using *co1* gene. Conditional on obtaining a statistically significant result from the KW-test, we then performed an exact test for increasing trend using the Jonckheere-Terpstra test (JT-test). The JT-test is similar to KW-test, but allows testing of the specific alternative hypothesis of a trend (decrease or increase) in the response variable amongst the groups being tested. The JT-test is therefore more powerful than the KW-test when dealing with ordinal variables, as is the case of Anthromes. We tested the hypothesis that there is an increase in genetic diversity, measured as  $\hat{\Pi}$ , from anthrome “Dense settlements” towards anthrome “Wildlands”, ordered as follows: “Dense settlements”, “Villages”, “Croplands”, “Rangelands”, “Forested” and “Wildlands”. KW and JT tests were run in the R package “*clinfun*” (51). Results show that genetic diversity differs among anthromes in amphibians (KW-test,  $\chi^2 = 15.271$ ,  $df = 5$ ,  $P = 0.009$ ) and mammals using *co1* gene (KW-test,  $\chi^2 = 22.437$ ,  $df = 5$ ,  $P < 0.001$ ) but not in mammals using *cytb* gene (KW-test,  $\chi^2 = 9.696$ ,  $df = 5$ ,  $P = 0.084$ ). Furthermore JT-tests show that there is an increase in genetic diversity from “Dense settlements” towards “Wildlands” in amphibians (JT-test,  $JT_{statistic} = 55006$ ,  $P = 0.004$ ) and in mammals using *co1* gene (JT-test,  $JT_{statistic} = 352761$ ,  $P = 0.025$ ).

### Hierarchical regression analysis

To estimate the independent effect of several predictor variables on genetic diversity we used a hierarchical regression analysis (52) as implemented in the “*hier.part*” R package (53). This regression approach allows estimating the independent effect (IE) values of each predictor variable on the spatial distribution of genetic diversity across grid cells. In other words, all possible models in a multiple regression setting are jointly considered to attempt to identify the most likely causal factors. Hierarchical partitioning includes the calculation of the incremental improvement (i.e. increased  $r^2$ ) in models by the addition of a given variable and these are averaged over all combinations in which that variable occurs to estimate the independent effects of each variable. We estimated the independent effects of nine predictor variables to explain the spatial variation of GD across grid cells. The predictor variables used are: latitude, longitude, number of base pairs and six other variables representing the area occupied by each anthrome category (“Dense settlements”, “Villages”, “Croplands”, “Forested” and “Wildlands”) within each grid cell. We ran all possible regression models independently for mammals and amphibians (512 models in each case) and estimated the total variance explained by the full model and the independent effect of each predictor variable. Moreover, we randomized 100 times the values in each independent variable and then ran the 512 regression models again to obtain 100 simulated values of IE. This randomization procedure allows us to estimate the 95th percentile values of IE values for the randomized dataset and the statistical significance of the observed IE values of each variable. Results show that in mammals both latitude and anthrome “Rangelands” have significant independent effects on GD, with latitude accounting for the largest amount of variance in GD (46.2%) followed by the area of “Rangelands” (33.3%) (Table S5). In amphibians, the area of anthromes “Forested” and “Wildlands” show the highest independent effects on GD (50.7% and 12.1%, respectively; Table S5).

### Choosing grid cell grain size for spatial analyses of genetic diversity

We have shown above that estimates of  $\hat{\Pi}$  deriving from species with very few sequences in a grid cell are associated with high uncertainty. On the other hand, the number of genetic sequences from each species that are included in the calculations of  $\hat{\Pi}$  depend on the area of the grid cell chosen. Therefore, in order to minimize the uncertainty in the calculations of  $\hat{\Pi}$  it was important to chose a grid cell area that maximizes the number of sequences from each species that are included in the calculations of  $\hat{\Pi}$  and minimizes the difference in the number of sequences between species ( $\sigma^2$ ). In order to identify the grid cell area that satisfies the above conditions we projected all georeferenced sequences into an equal area projection (Behrmann) and overlaid them with four nested equal area grids with grain sizes of c. 96.5 km, 193 km, 385.9 km and 771.9 km, representing 1°, 2°, 4 and 8° longitude at 30°N (the point of no-distortion in a Behrmann projection), respectively. For each grid cell grain size we calculated the number of sequences per species in each grid cell. Results show that the median number of sequences per species is highest in the coarser grain sizes (385.9 km and 771.9 km) but increasing the grain size from 385.9 km to 771.9 km leads to a substantial increase in variance (85% increase;  $\sigma^2_{385.9} = 386.0$ ,  $\sigma^2_{771.9} = 712.4$ ) (fig. S15). Another important aspect in chosing the grain

size concerns the error associated with our georeferencing method. Results from our quality control of georeferenced sequences show that the average distance between localities assigned using GeoNames.org and localities originally uploaded as annotations (see section “Quality control of georeferenced sequences”) is 216 km, and therefore in order to minimize the effects of georefencing errors, a coarser grain size was selected. To assess the sensitivity of our results to the choice of grain size we estimated the distribution of genetic diversity using a coarser grain (771.9 km). Patterns of the distribution of genetic diversity remain unchanged (figs. S17 and S18).

**Table S1**

Total number of sequences and species with data available for each locus. The proportion of sequences with geographic coordinates and the proportion of species for which we have at least one sequence with geographic coordinates are indicated in parentheses.

		<b>Mammals</b>	<b>Amphibians</b>	<b>Combined</b>
all mitochondrial loci	<b>Sequences</b>	236056 (30%)	82692 (38%)	318748 (32%)
	<b>Species</b>	3930 (55%)	3195 (64%)	7125 (60%)
<i>cytb</i>	<b>Sequences</b>	61580 (24%)	24823 (38%)	86403 (28%)
	<b>Species</b>	3139 (36%)	1536 (56%)	4675 (43%)
<i>col</i>	<b>Sequences</b>	32887 (69%)		
	<b>Species</b>	1763 (55%)		

**Table S2**

Total sequences available (N) and number of georeferenced sequences (N georeferenced) per locus for terrestrial mammals and amphibians. The percentage of sequences per locus relative to the total amount of sequences available for all loci is shown in parentheses.

<b>locus</b>	<b>N</b>	<b>N georeferenced</b>
Mammals:		
all loci	236056	69666
<i>col</i>	44893 (19%)	31739 (46%)
<i>cytb</i>	61580 (26%)	14959 (22%)
<i>NADH</i>	14717 (6%)	2255 (3%)
Amphibians:		
all loci	82692	31491
<i>cytb</i>	24823 (11%)	9520 (30%)
<i>NADH</i>	13577 (16%)	4335 (14%)
<i>col, co2 and co3</i>	10843 (13%)	4386 (14%)

**Table S3**

List of domesticated, managed, laboratory reared or extinct (\*) species excluded from analyses.

<i>Alces alces</i>	<i>Moschus moschiferus</i>
<i>Alces americanus</i>	<i>Mus musculus</i>
<i>Axis axis</i>	<i>Oryctolagus cuniculus</i>
<i>Axis porcinus</i>	<i>Ovis aries</i>
<i>Bison bonasus</i>	<i>Rangifer tarandus</i>
<i>Bison bison</i>	<i>Rattus norvegicus</i>
<i>Bos frontalis</i>	<i>Sus scrofa</i>
<i>Bos grunniens</i>	<i>Vicugna pacos</i>
<i>Bos indicus</i>	<i>Xenopus laevis</i>
<i>Bos javanicus</i>	<i>Bison priscus</i> *
<i>Bos taurus</i>	<i>Bos primigenius</i> *
<i>Bubalus bubalis</i>	<i>Caloprymnus campestris</i> *
<i>Camelus bactrianus</i>	<i>Chaeropus ecaudatus</i> *
<i>Camelus dromedarius</i>	<i>Dusicyon australis</i> *
<i>Canis familiaris</i>	<i>Gazella saudiya</i> *
<i>Canis lupus</i>	<i>Hippotragus leucophaeus</i> *
<i>Canis lycaon</i>	<i>Homo heidelbergensis</i> *
<i>Capra hircus</i>	<i>Hydrodamalis gigas</i> *
<i>Cavia porcellus</i>	<i>Malpaisomys insularis</i> *
<i>Cervus elaphus</i>	<i>Mammut americanum</i> *
<i>Cervus hortulorum</i>	<i>Mammuthus columbi</i> *
<i>Cervus nippon</i>	<i>Mammuthus primigenius</i> *
<i>Cervus timorensis</i>	<i>Megaladapis edwardsi</i> *
<i>Cervus unicolor</i>	<i>Monachus tropicalis</i> *
<i>Dama dama</i>	<i>Palaeopropithecus ingens</i> *
<i>Elaphurus davidianus</i>	<i>Perameles eremiana</i> *
<i>Equus asinus</i>	<i>Potorous platyops</i> *
<i>Equus caballus</i>	<i>Pteropus tokudaee</i> *
<i>Felis catus</i>	<i>Rattus macleari</i> *
<i>Felis sivestris</i>	<i>Thylacinus cynocephalus</i> *
<i>Lama glama</i>	<i>Ursus deningeri</i> *
<i>Lama guanicoe</i>	<i>Ursus spelaeus</i> *
<i>Lama pacos</i>	<i>Zalophus japonicus</i> *
<i>Lama vicugna</i>	

**Table S4**

Taxonomic discrepancies between IUCN, GenBank and BOLD databases. In these cases, species were treated as one single species, following the taxonomic nomenclature of IUCN.

IUCN	GenBank/BOLD
<i>Spalax ehrenbergi</i>	<i>Nannospalax ehrenbergi</i> <i>Nannospalax galili</i> <i>Nannospalax golani</i> <i>Nannospalax judaei</i> <i>Spalax carmeli</i>
<i>Rhinolophus pusillus</i>	<i>Rhinolophus cornutus</i> <i>Rhinolophus monoceros</i> <i>Rhinolophus pumilus</i>
<i>Necromys urichi</i>	<i>Akodon urichi</i> <i>Bolomys urichi</i>
<i>Pseudoeurycea bellii</i>	<i>Isthmura bellii</i> <i>Isthmura sierraoccidentalis</i>
<i>Ranitomeya sirensis</i>	<i>Ranitomeya biolat</i> <i>Ranitomeya lamasi</i>
<i>Osteocephalus buckleyi</i>	<i>Osteocephalus cannatellai</i> <i>Osteocephalus vilmae</i>
<i>Lithobates berlandieri</i>	<i>Rana berlandieri</i> <i>Rana brownorum</i>
<i>Speleomantes italicus</i>	<i>Hydromantes ambrosii</i> <i>Hydromantes italicus</i>

**Table S5**

Independent effects of explanatory variables and their significance on genetic diversity. The independent effects (IE) values correspond to the percentage of the explained variance accounted for by each explanatory variable as calculated by hierarchical partitioning. Observed IE values for each variable, Z-scores for the generated distribution of randomized IE and an indication of statistical significance. Z-scores are calculated as (observed - mean(randomizations))/sd(randomizations), nd statistical significance (\*) is based on upper 0.95 confidence limit ( $Z \geq 1.65$ ). Variables are ordered by effect size.

	<b>IE (%)</b>	<b>Observed IE</b>	<b>Z score</b>
<b>Mammals</b>			
Latitude	46.24	0.00	9.39*
Rangelands	33.26	0.00	7.22*
Villages	7.12	0.00	1.16
Dense settlements	6.27	0.00	1.22
Wildlands	2.52	0.00	0.02
Longitude	1.80	0.00	-0.27
Forested	1.67	0.00	-0.38
Croplands	0.92	0.00	-0.42
Base pairs	0.21	0.00	-0.76
<b>Amphibians</b>			
Forested	50.65	0.02	37.51*
Wildlands	12.12	0.00	4.12*
Base pairs	11.23	0.00	9.08*
Longitude	8.88	0.00	3.11*
Latitude	5.66	0.00	2*
Croplands	5.56	0.00	2.59*
Rangelands	2.52	0.00	1.06
Villages	1.69	0.00	0.07
Dense settlements	1.69	0.00	0.12

## **Additional file S1**

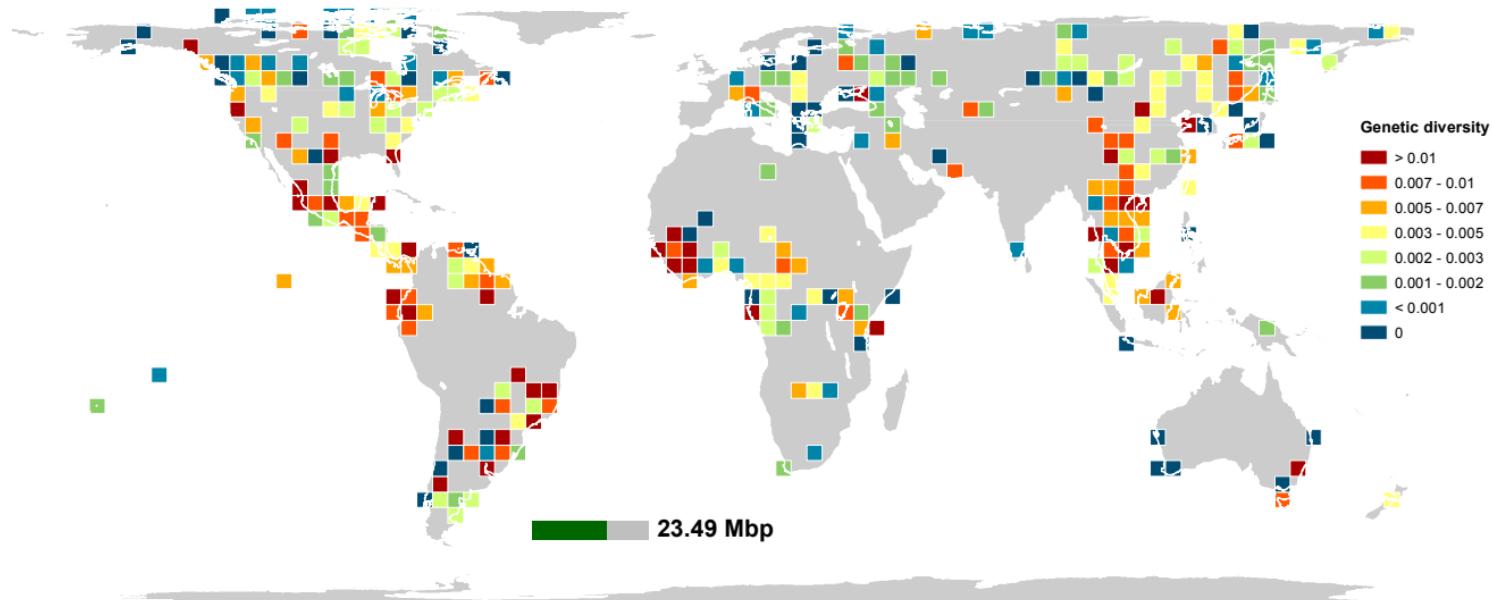
### **Analysis source code**

ZIP archive of source code used for all analyses described in the manuscript. The archive contains source files written for *julia* and *R*, and includes a README file detailing how to create all figures and analyses presented in the main text and supplementary information.

**Additional file S2**

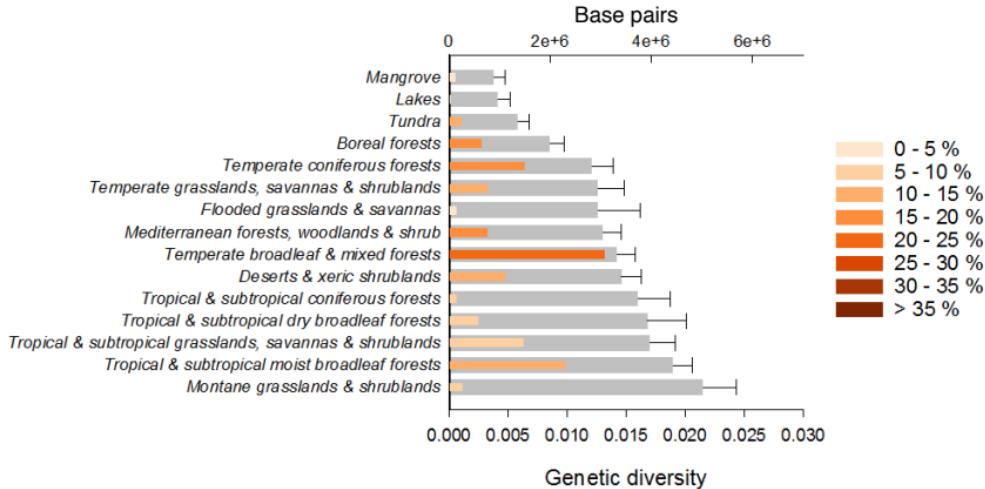
Document explaining how to use the web-based tool iMapGenes that can be accessed at  
<http://macroecology.ku.dk/resources/imapgenes>.

## Mammals *co1*



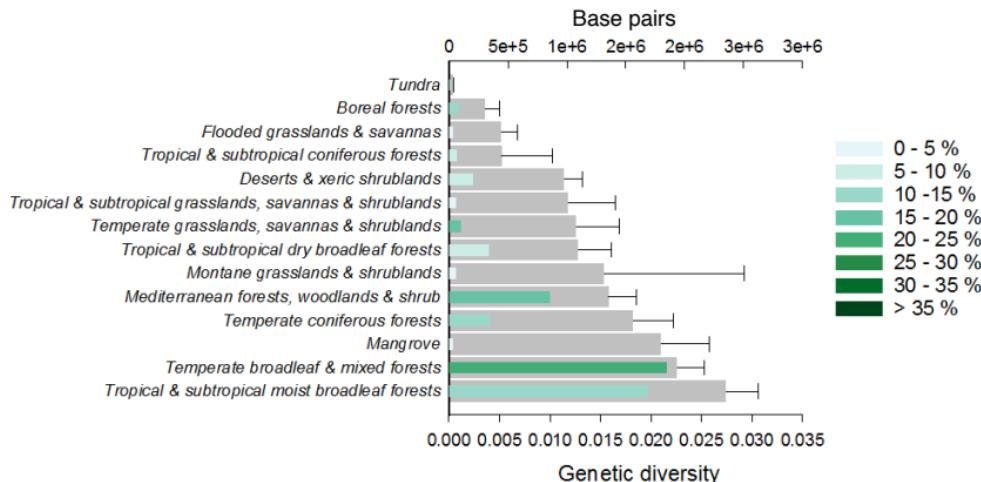
**Fig. S1 Global distribution of genetic diversity based on *cytochrome oxidase 1* (*co1*), for terrestrial mammals.**

Average number of mutations per base pair for *co1* in mammals. Colors represent quantiles (8 quantiles). The gray bar below the map represents the total number of *co1* base pairs retrieved from GenBank and BOLD, and the green bar represents the number of georeferenced base pairs that were used to estimate the global distribution of genetic diversity.



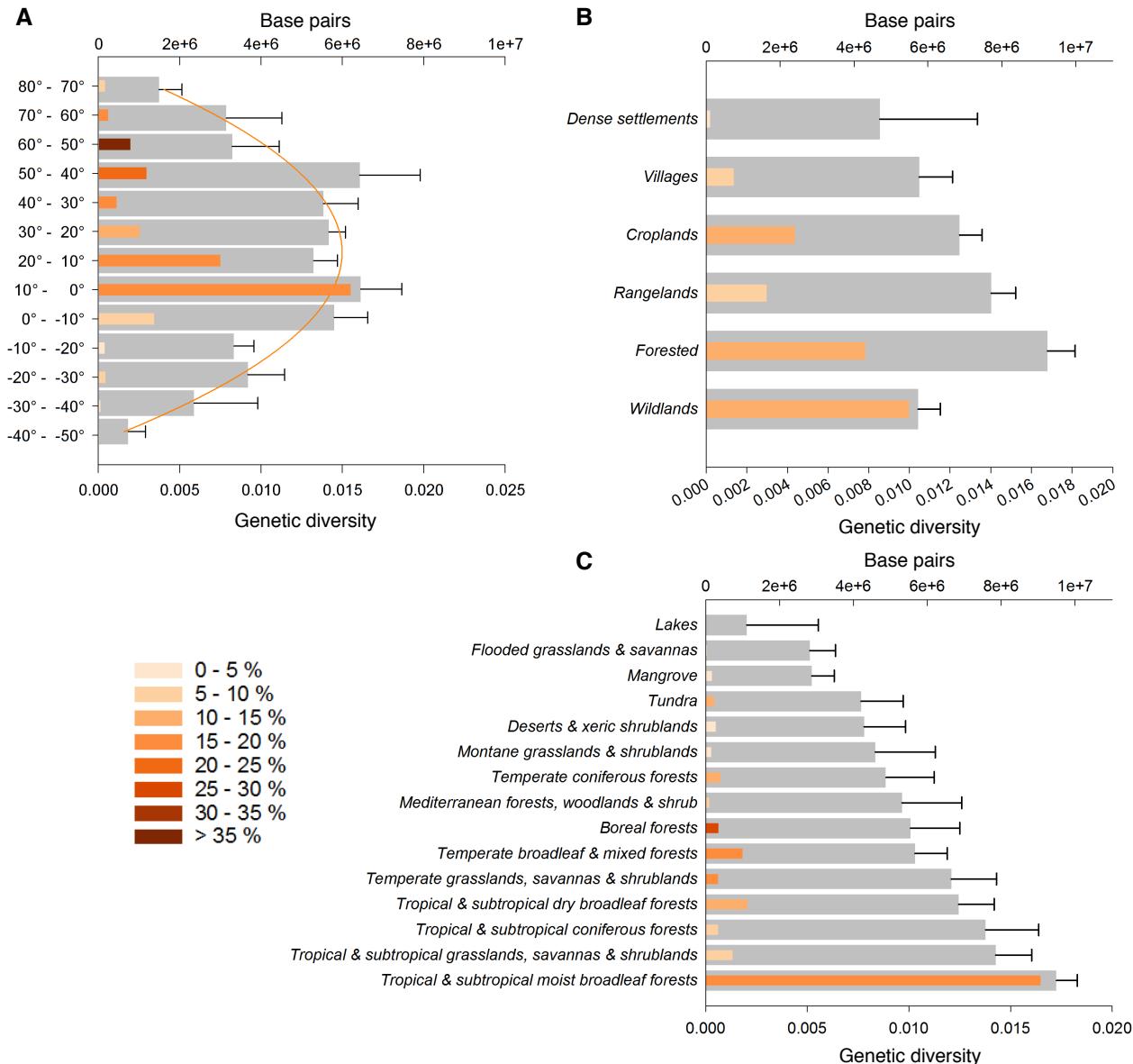
**Fig. S2 Distribution of mammalian genetic diversity across biomes, for *cytochrome b* (*cytb*).**

Gray bars (lower x-axis) indicate the average number of mutations per base pair for *cytb* across species of terrestrial mammals in each biome. The length of colored bars (upper x-axis) indicates the number of *cytb* base pairs used in the calculation of GD, while the depth of color shading denotes the percentage of species used in the calculation of GD relative to the species richness of each category. The distribution of GD across biomes is highly concordant with that observed for mammals using *co1* (Fig. S4C), with tropical biomes showing higher GD. Error bars represent standard deviations from the mean of the bootstrapping analysis.



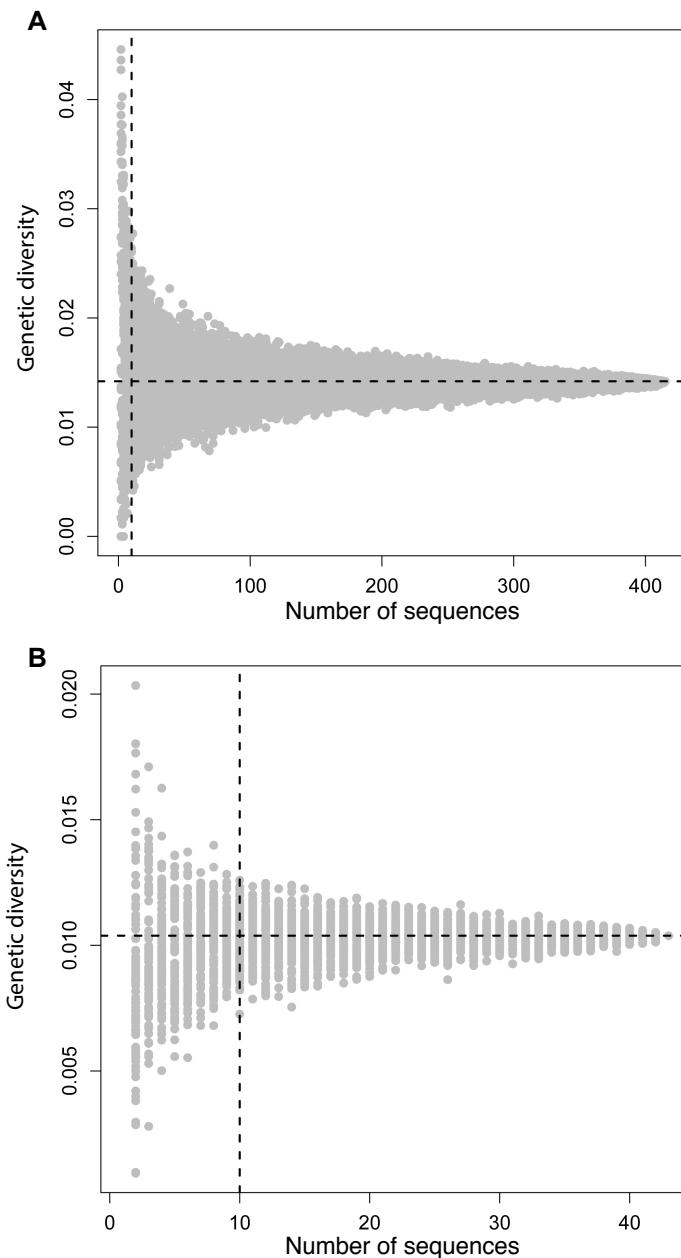
### Fig. S3 Distribution of amphibian genetic diversity across biomes, for *cytochrome b* (*cytb*).

Gray bars (lower x-axis) indicate the average number of mutations per base pair for *cytb* across species of amphibians in each biome. The length of colored bars (upper x-axis) indicates the number of *cytb* base pairs used in the calculation of GD, while the depth of color shading denotes the percentage of species used in the calculation of GD relative to the species richness of each category. The distribution of amphibian GD at the biome level shows higher GD for tropical and subtropical broadleaf forests, as observed for mammals (Fig. S2 and S4C); however, other tropical biomes have lower GD than their temperate counterparts. Error bars represent standard deviations from the mean of the bootstrapping analysis.



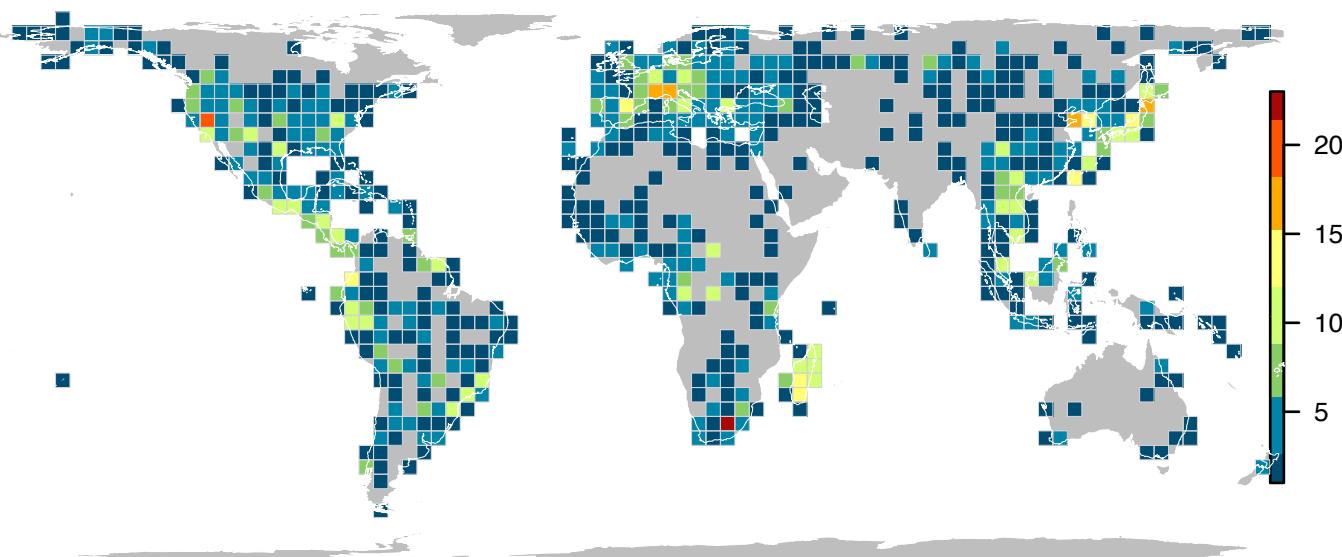
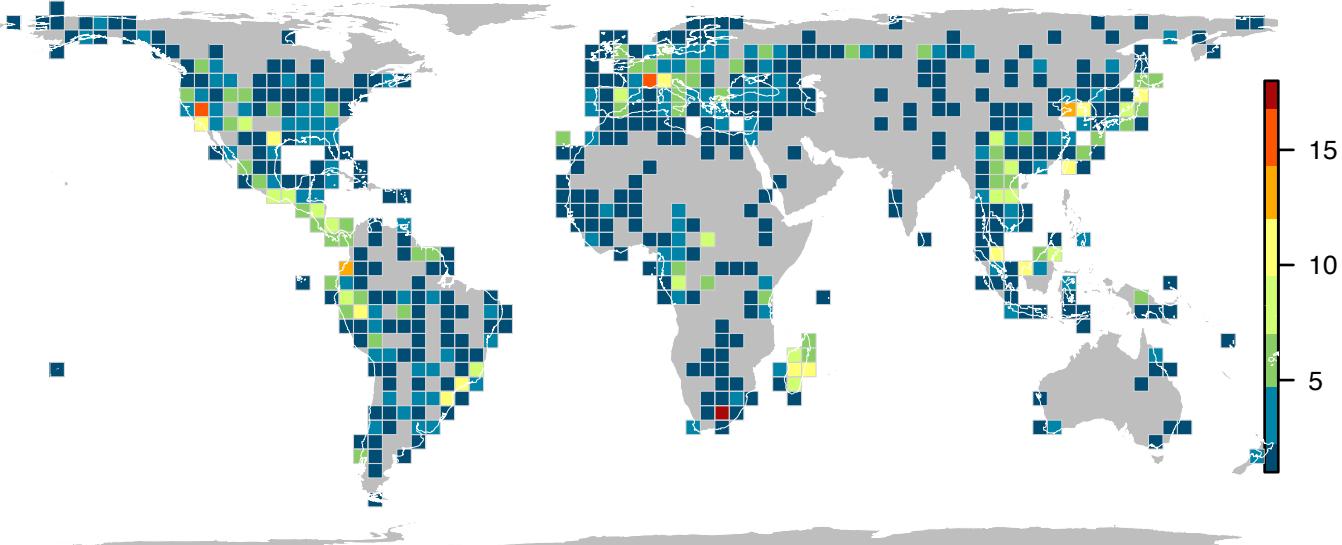
**Fig. S4 Distribution of mammalian genetic diversity across latitudes, anthromes and biomes, for cytochrome oxidase 1 (*co1*).**

Gray bars (lower x-axis) indicate the average number of mutations per base pair for *co1* across species of terrestrial mammals in each latitudinal band (A), anthrome (B) and biome (C). The length of colored bars (upper x-axis) indicates the number of *co1* base pairs used in the calculation of GD, and the depth of color shading denotes the percentage of species used in the calculation of GD relative to the species richness of each category. The distribution of GD per latitudinal band is highly congruent with that observed for both mammals using *cytb* (Fig. 3A,  $r = 0.79$ ,  $P = 0.001$ ) and amphibians (Fig. 3C,  $r = 0.66$ ,  $P = 0.01$ ). The quadratic polynomial model shown in (A) indicates a peak of GD in the tropics (pseudo  $r^2 = 0.89$ ). The distribution of GD per anthrome differs slightly from what is observed for mammals using *cytb*, where villages and croplands show the highest levels of GD (Fig. 3B). At the biome level, the distribution of GD is highly concordant with what is observed for mammals using *cytb* (Fig. S2), with tropical biomes showing higher levels of GD. Error bars represent standard deviations from the mean of the bootstrapping analysis.



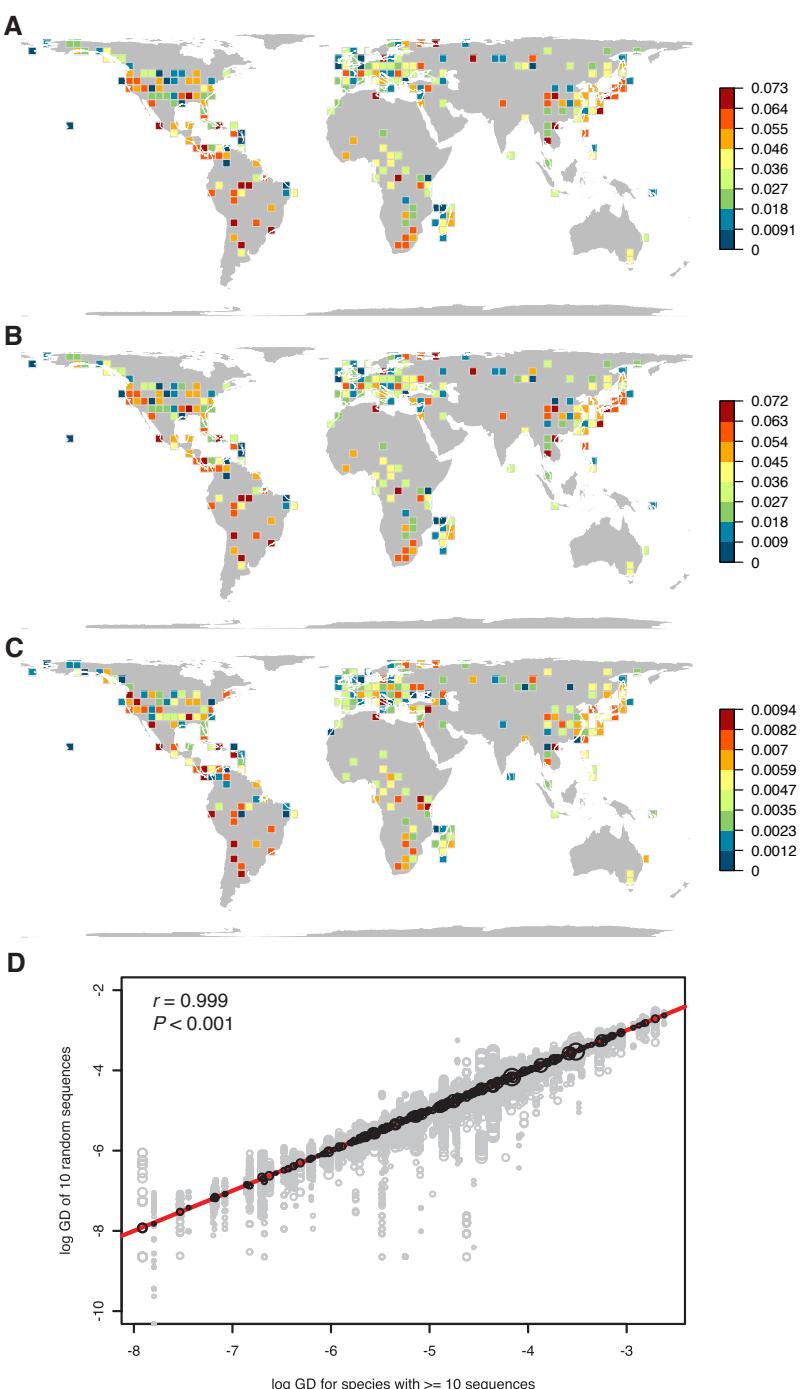
**Fig. S5 Effect of sequences availability on nucleotide diversity per site.**

Genetic diversity (GD), measured as nucleotide diversity per site, for *Myodes glareolus* (**A**) and *Rana sauteri* (**B**) estimated from 100 random subsamples of each alignment at each sample size between 2 and N total number of sequences. Although the expected value of GD is not dependent on the number of sequences, the standard error is.

**A****B**

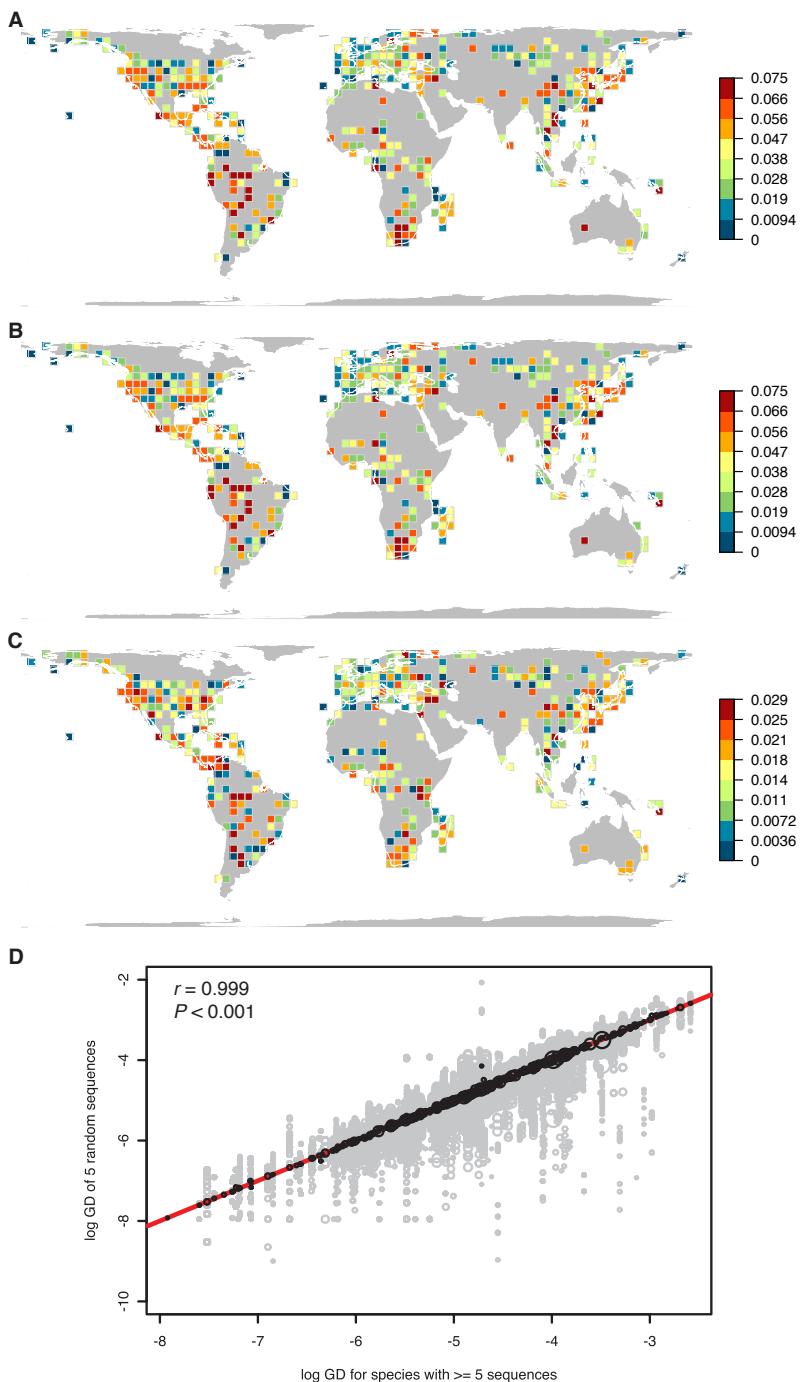
**Fig. S6 Geographic distribution of species with small sample sizes.**

Species richness based on species with fewer than 10 (A) and fewer than 5 (B) sequences per equal area grid cell. There is no consistent spatial bias in the locations of these species.



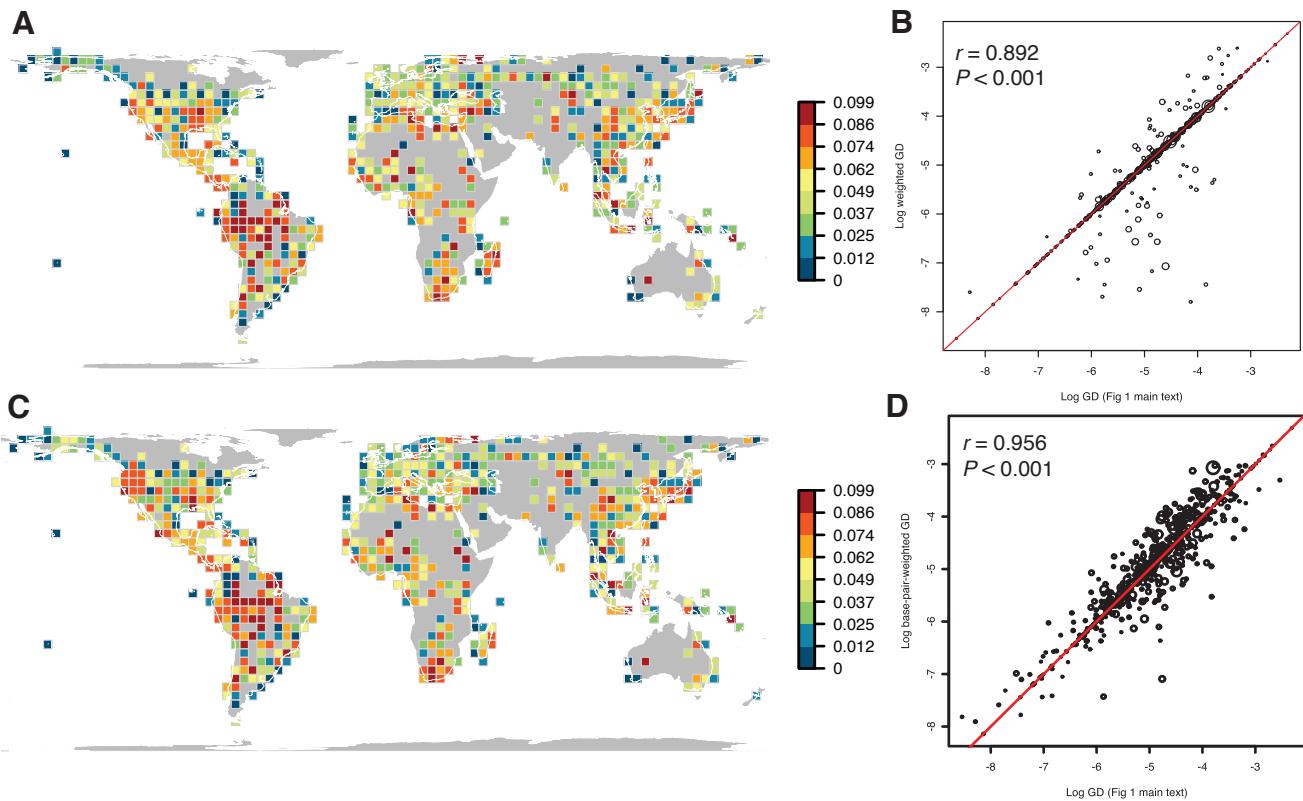
**Fig. S7 Effect of sequence availability on genetic diversity.**

**(A)** Spatial distribution pattern of GD (similar to those in Fig. 1A) calculated for those species with more than 10 sequences per grid cell. For each grid cell we identify all species that have  $<10$  sequences, and exclude those sequences from the GD calculations of that grid cell. Sequences were excluded on a per-grid cell basis, such that an individual species might ‘lose’ some grid cells but not disappear from the analysis. The overall distribution of GD shown in Fig. 1A is unchanged. **(B)** Spatial distribution patterns of the mean and standard deviation **(C)** of GD estimated by the rarefaction analysis. **(D)** Relationship between mean GD of the rarefaction analysis **(B)** and GD in **(A)**. Gray circles represent the GD values of the 100 random draws from the rarefaction analysis, and black circles represent the mean values of these 100 repetitions. Circle size indicates the total number of base pairs available per grid cell. The red line is the 1:1 line. Mean GD from the rarefaction analysis was consistent with GD for species with  $>10$  sequences in each grid cell (Pearson correlation  $r = 0.999$ ,  $P < 0.001$ ). Correlation significance was determined using Dutilleul’s (47) spatially corrected degrees of freedom.



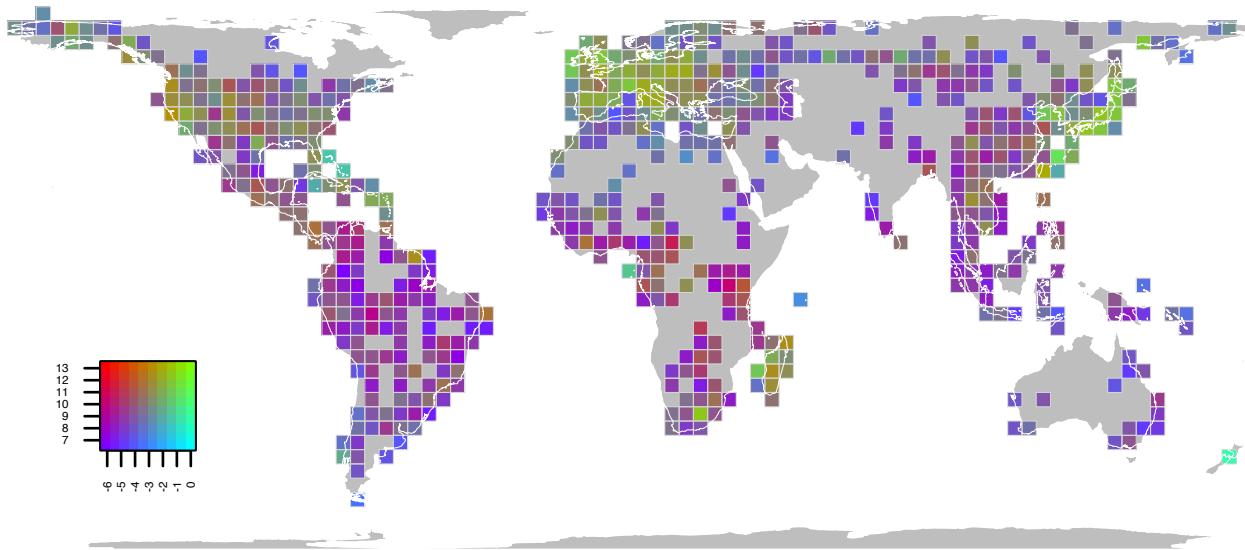
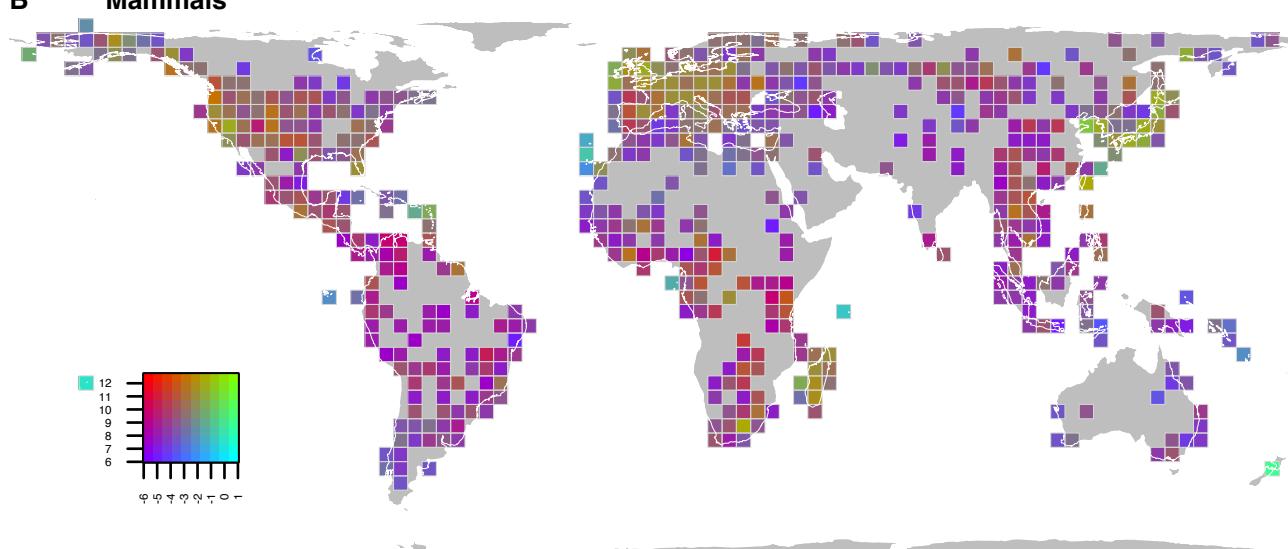
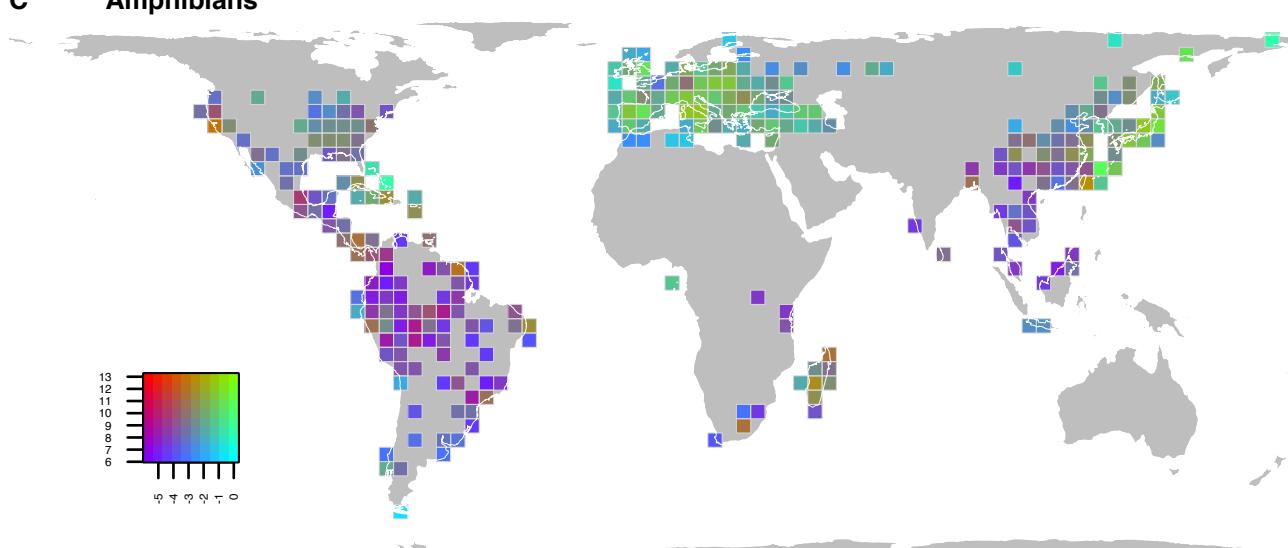
**Fig. S8 Effect of sequence availability on genetic diversity.**

**(A)** Spatial distribution pattern of GD (similar to those in Fig. 1A) calculated for those species with more than 5 sequences per grid cell. For each grid cell we identify all species that have  $<5$  sequences, and exclude those sequences from the GD calculations of that grid cell. Sequences were excluded on a per-grid cell basis, such that an individual species might ‘lose’ some grid cells but not disappear from the analysis. The overall distribution of GD shown in Fig. 1A is unchanged. **(B)** Spatial distribution patterns of the mean and standard deviation **(C)** of GD estimated by the rarefaction analysis. **(D)** Relationship between mean GD of the rarefaction analysis **(B)** and GD in **(A)**. Gray circles represent the GD values of the 100 random draws from the rarefaction analysis, and black circles represent the mean values of these 100 repetitions. Circle size indicates the total number of base pairs available per grid cell. The red line is the 1:1 line. Mean GD from the rarefaction analysis was consistent with GD for species with  $>5$  sequences in each grid cell (Pearson correlation  $r = 0.999$ ,  $P < 0.001$ ). Correlation significance was determined using Dutilleul’s (47) spatially corrected degrees of freedom.

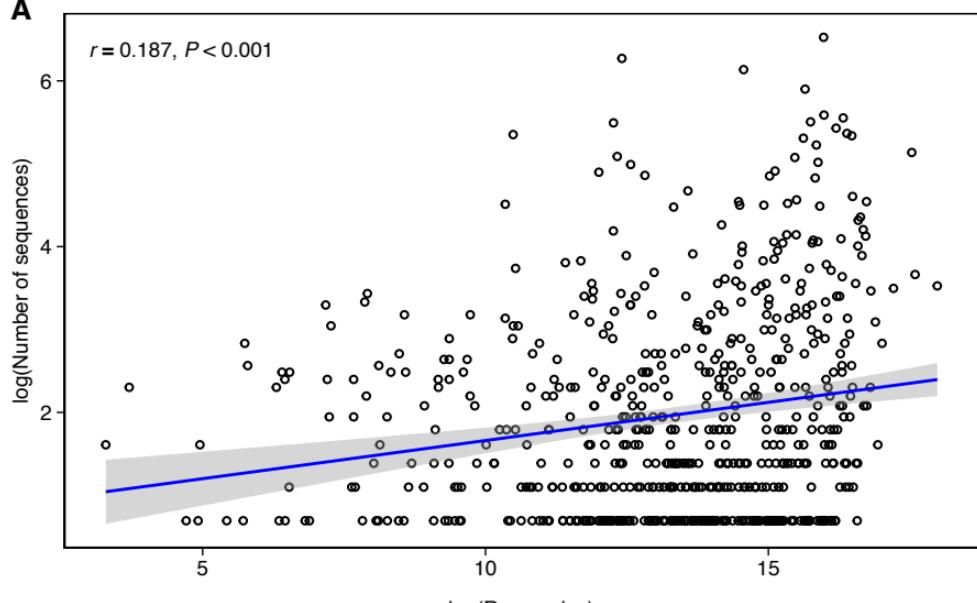
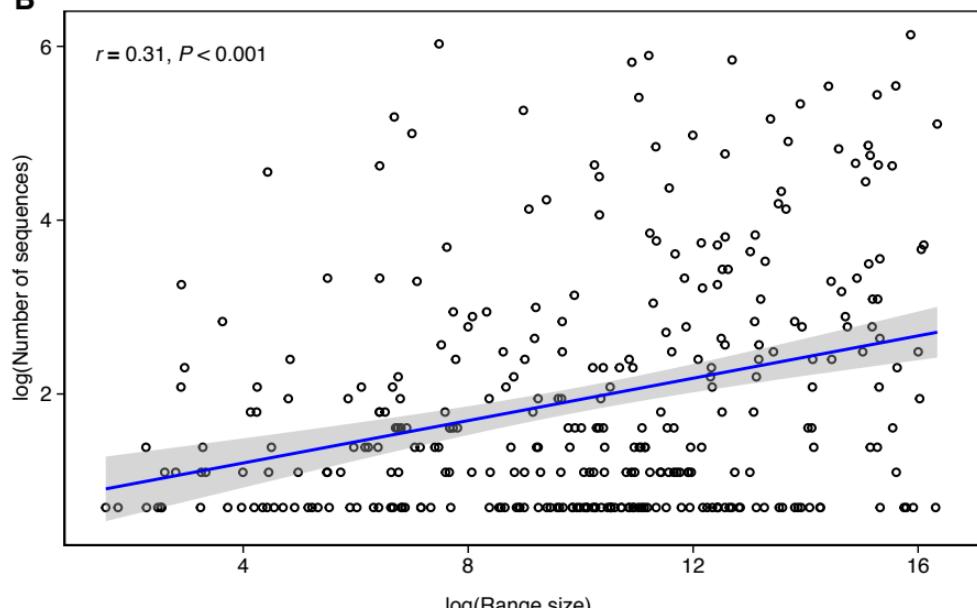


**Fig. S9 Spatial effect of weighting Genetic Diversity by sampling intensity.**

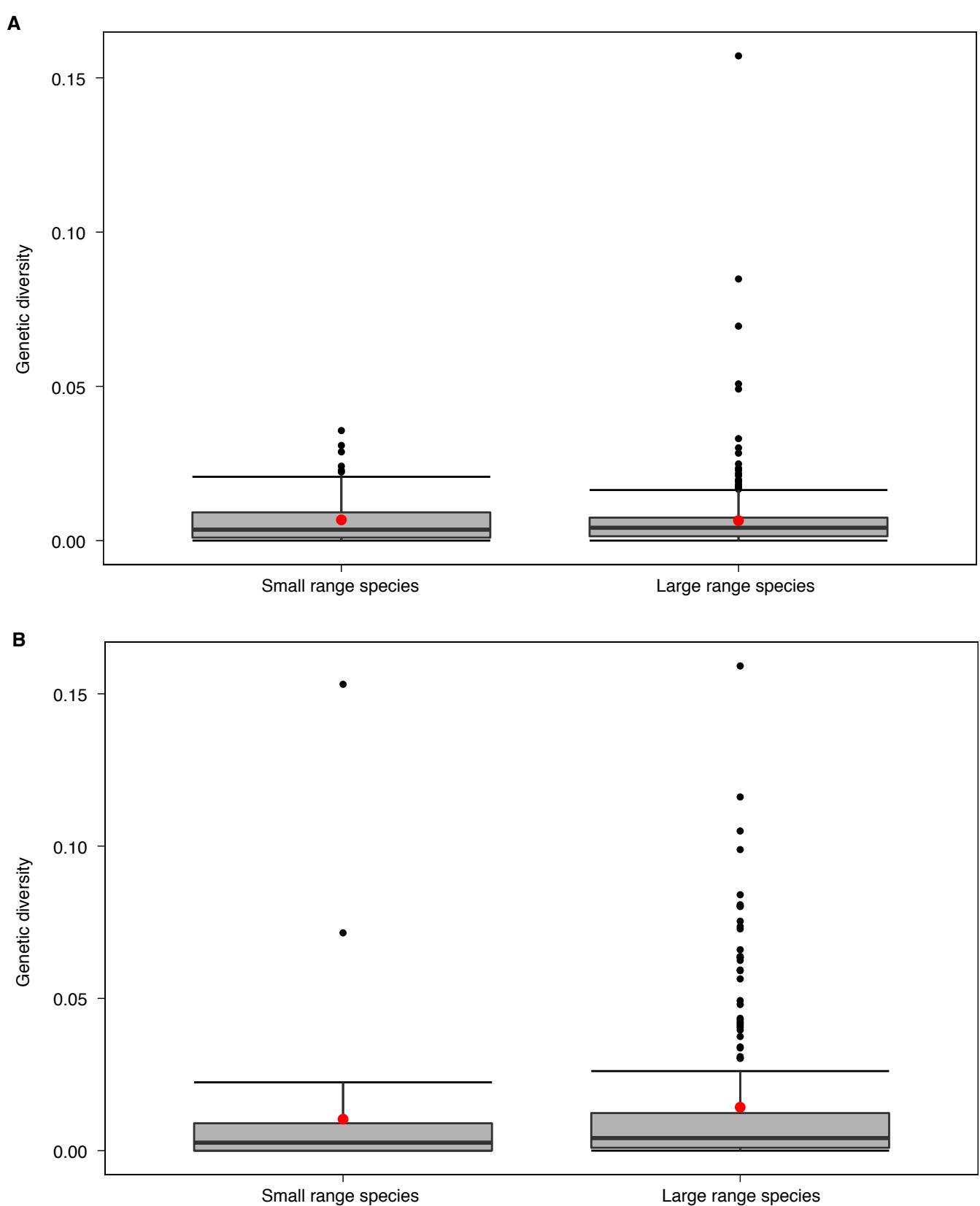
Spatial distribution of GD weighted by the number of sequence pairs within each equal area grid cell (grain size 385.9 km) (**A**) and its relationship with unweighted GD as presented in Fig. 1A (**B**). The weighted GD is very highly concordant with the GD presented in Fig. 1A (Pearson correlation  $r = 0.892$ ,  $P < 0.001$ ). Spatial distribution of GD weighted by the number of base pairs within each equal area grid cell (grain size 385.9 km) (**C**) and its relationship with unweighted GD as presented in Fig. 1A (**D**). The weighted GD is very highly concordant with the GD presented in Fig. 1A (Pearson correlation  $r = 0.956$ ,  $P < 0.001$ ). The red line in (B) and (D) is the 1:1 line. Correlation significance was determined using Dutilleul's (47) spatially corrected degrees of freedom.

**A Combined taxon****B Mammals****C Amphibians**

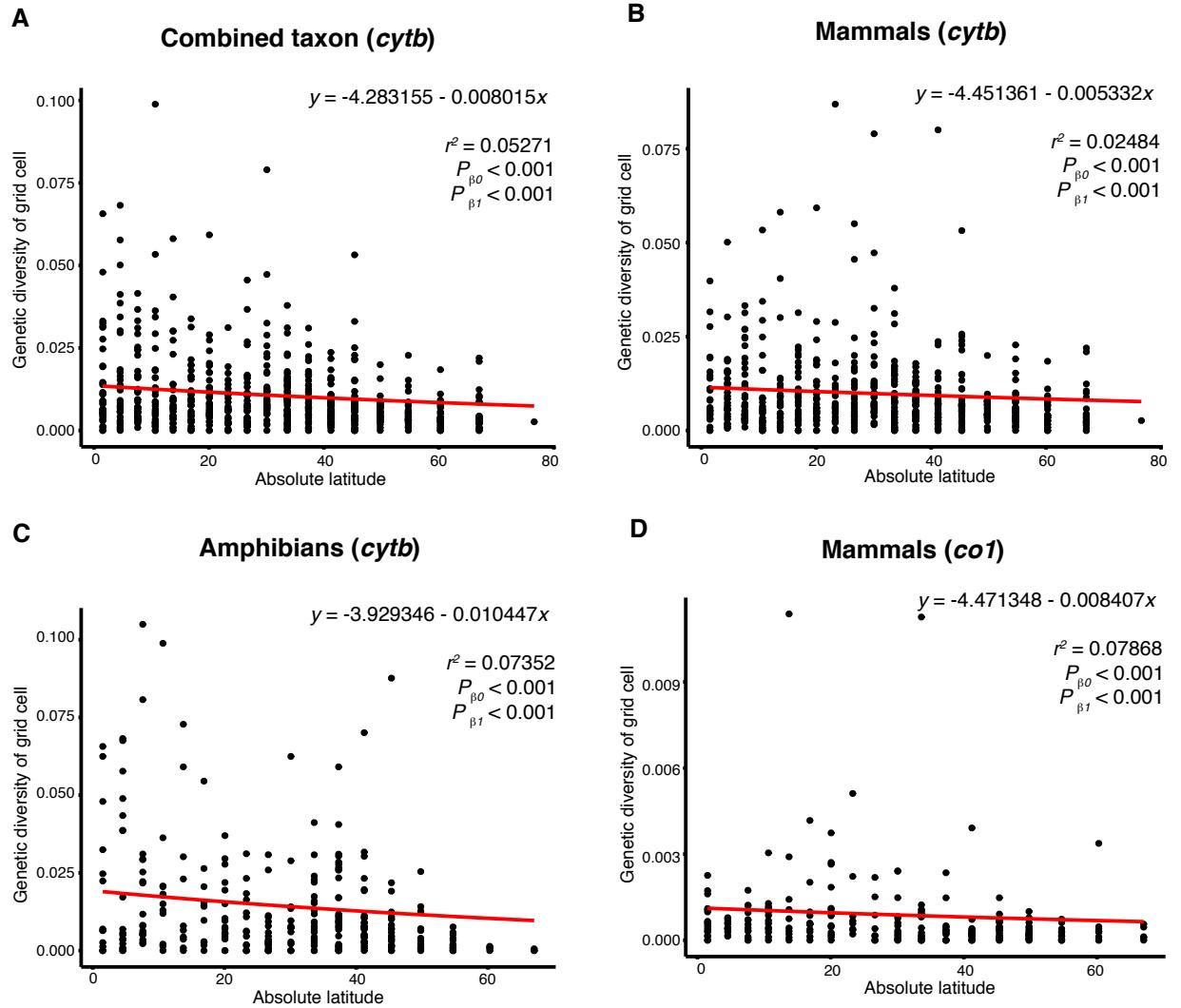
**Fig. S10 Global distribution of knowledge and ignorance using *cytochrome b* (*cytb*) data for mammals and amphibians together (A), mammals alone (B) and amphibians alone (C).** In the color scale (lower left corner of each map), the x-axis indicates the taxonomic coverage, defined by the percentage (log scale) of species known to inhabit each grid cell for which at least one georeferenced *cytb* sequence is available for that cell. The y-axis indicates sequence availability, i.e., the total number of *cytb* georeferenced base pairs (log scale) within each grid cell. In each map, violet colors indicate regions of low taxonomic coverage and relatively few georeferenced base pairs, while green colors indicate higher taxonomic coverage and higher numbers of base pairs. Light blue colors indicate high taxonomic coverage and few base pairs, while red colors indicate low taxonomic coverage and high numbers of base pairs.

**A****B**

**Fig. S11 Relationship between species range size and the number of georeferenced sequences available, for mammals (A) and amphibians (B).** The blue line represents the linear relationship with 95% confidence intervals in gray shade. Pearson correlation values ( $r$ ) and level of significance ( $P$ ) are shown in each graph.

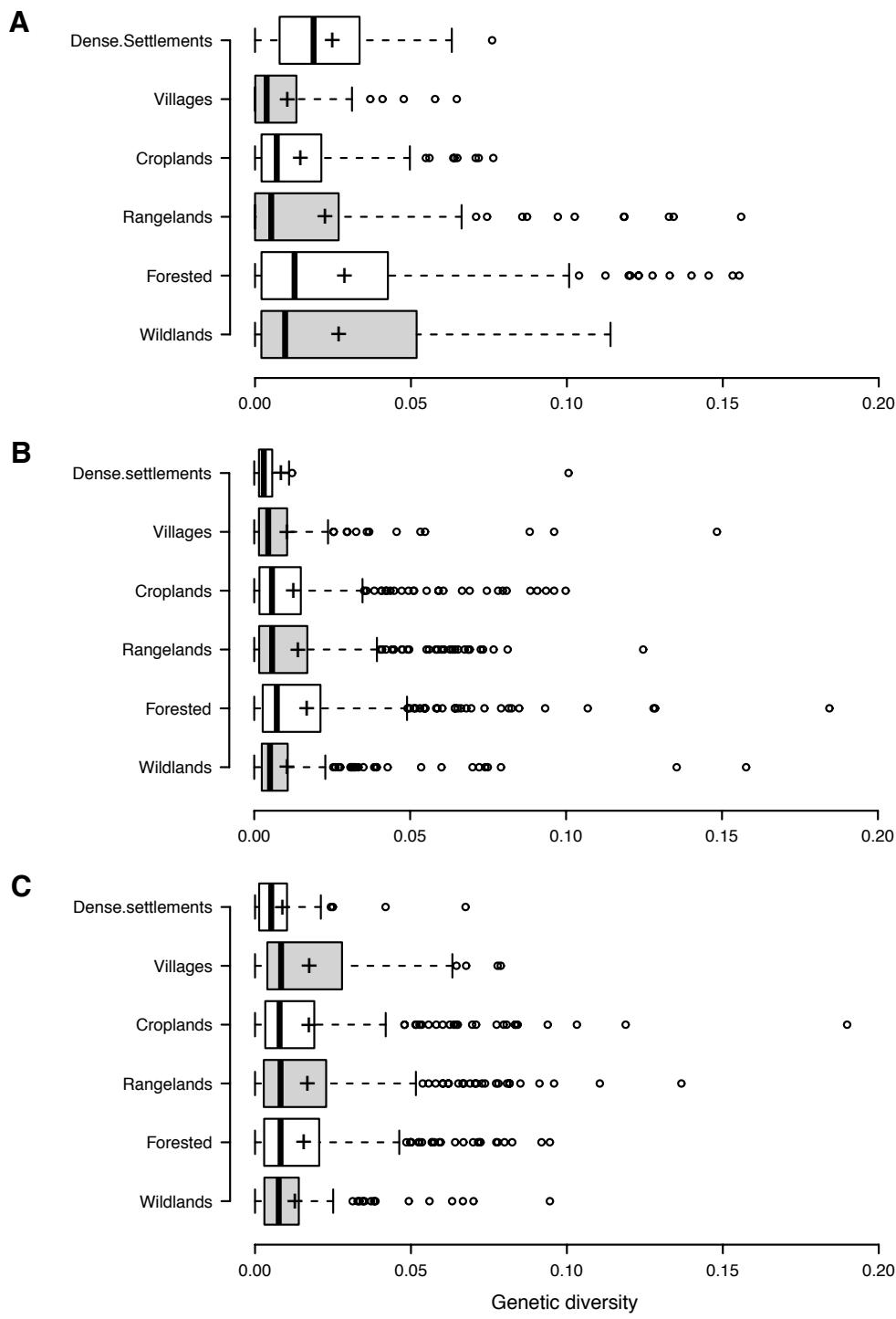


**Fig. S12 Genetic diversity for small range species (10<sup>th</sup> percentile in terms of range size) and large range species (90<sup>th</sup> percentile) of terrestrial mammals (A) and amphibians (B). Red dots represent the mean. There are no significant differences between groups (Mammals: Kruskal-Wallis  $\chi^2 = 0.0156$ ,  $P = 0.9007$ ; Amphibians: Kruskal-Wallis  $\chi^2 = 2.9837$ ,  $P = 0.08411$ ).**

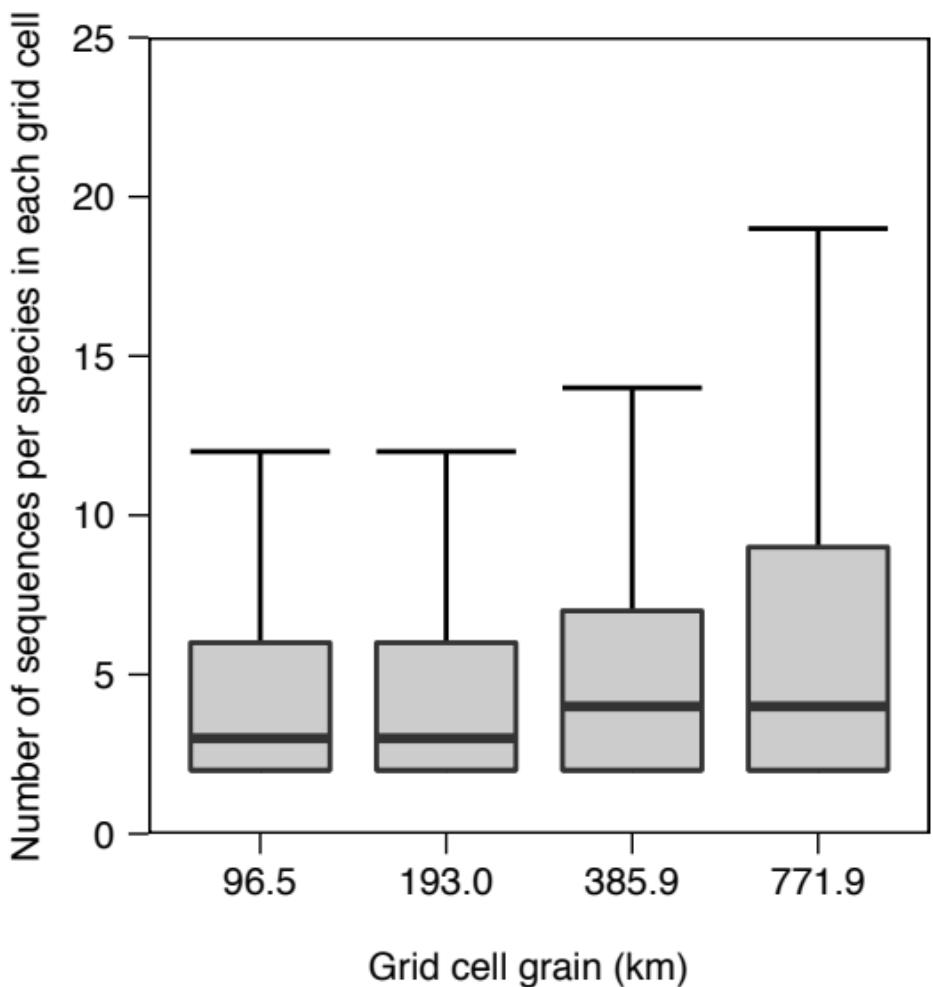


**Fig. S13 Relationship between latitude and genetic diversity.**

The scatterplots show genetic diversity (y-axis) plotted against absolute centroid latitude of each grid cell (x-axis). Regressions represent the predictions from beta regression linear models. For each regression we show the equation, the pseudo  $r^2$  value and the significance level ( $P$  values) for each term of the equation.

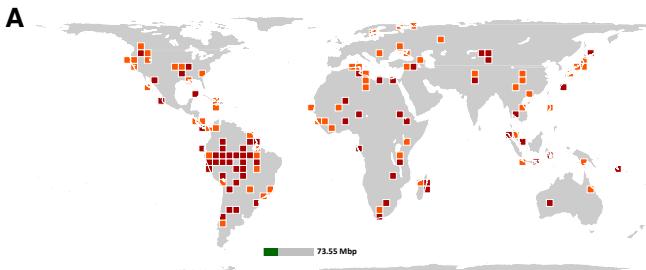


**Fig. S14 Relationship between anthrome type and genetic diversity in amphibians for *cytb* (A), terrestrial mammals for *co1* (B) and terrestrial mammals for *cytb* (C).** The boxplots show for each anthrome (y-axis) the distribution of genetic diversity values measured as nucleotide diversity per base pair for each species (x-axis). Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by open circles; crosses represent sample means as shown in Fig. 3 of the main text. Genetic diversity differs among anthromes in amphibians (A) (Kruskal-Wallis test,  $\chi^2 = 15.271$ , df = 5,  $P = 0.009$ ) and mammals using *co1* gene (B) (Kruskal-Wallis test,  $\chi^2 = 22.437$ , df = 5,  $P < 0.001$ ) but not in mammals using *cytb* gene (C) (Kruskal-Wallis test,  $\chi^2 = 9.696$ , df = 5,  $P = 0.084$ ). Furthermore Jonckheere-Terpstra tests show that there is an increase in genetic diversity from “Dense settlements” towards “Wildlands” in amphibians (Jonckheere-Terpstra test,  $JT_{statistic} = 55006$ ,  $P = 0.004$ ) and in mammals using *co1* gene (Jonckheere-Terpstra test,  $JT_{statistic} = 352761$ ,  $P = 0.025$ ).

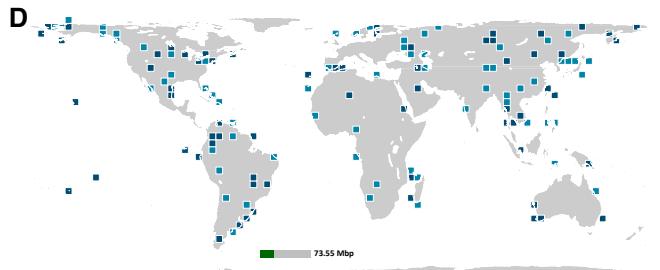
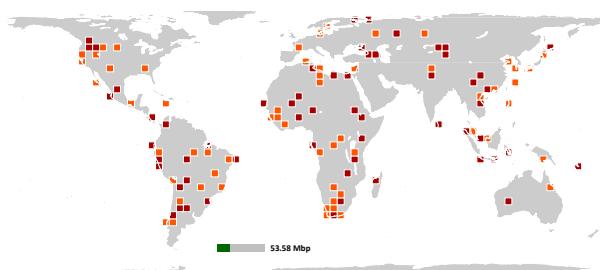
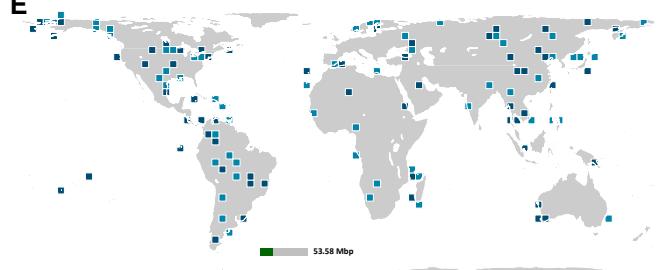
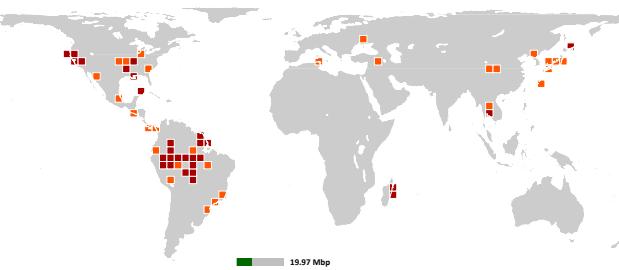
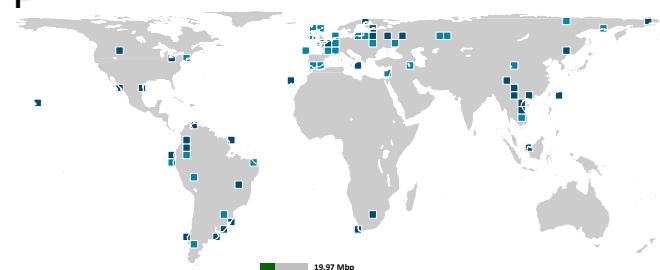


**Fig. S15 Number of sequences per species in each grid cell for each of the four nested equal area grain sizes (96.5 km, 193.0 km, 385.9 km and 771.9 km).** The median number of sequences per species in each grid cell is higher in the coarser grid cells (385.9 and 771.9), but increasing the grain size from 385.9 km to 771.9 km leads to a substantial increase in variance (85% increase;  $\sigma^2_{385.9} = 386.0$ ,  $\sigma^2_{771.9} = 712.4$ ).

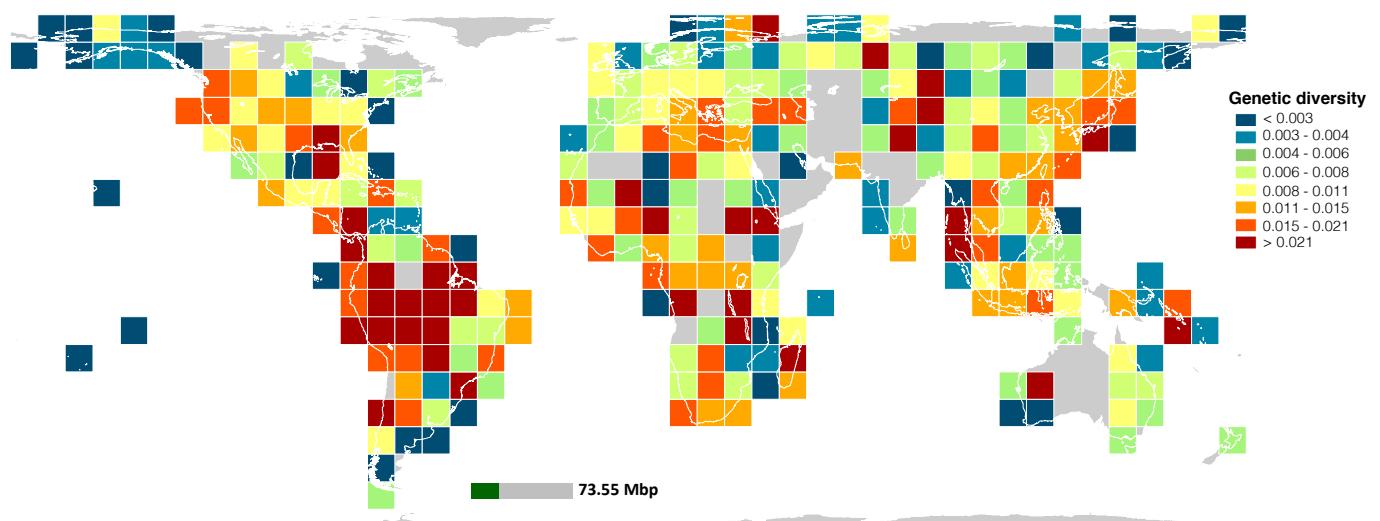
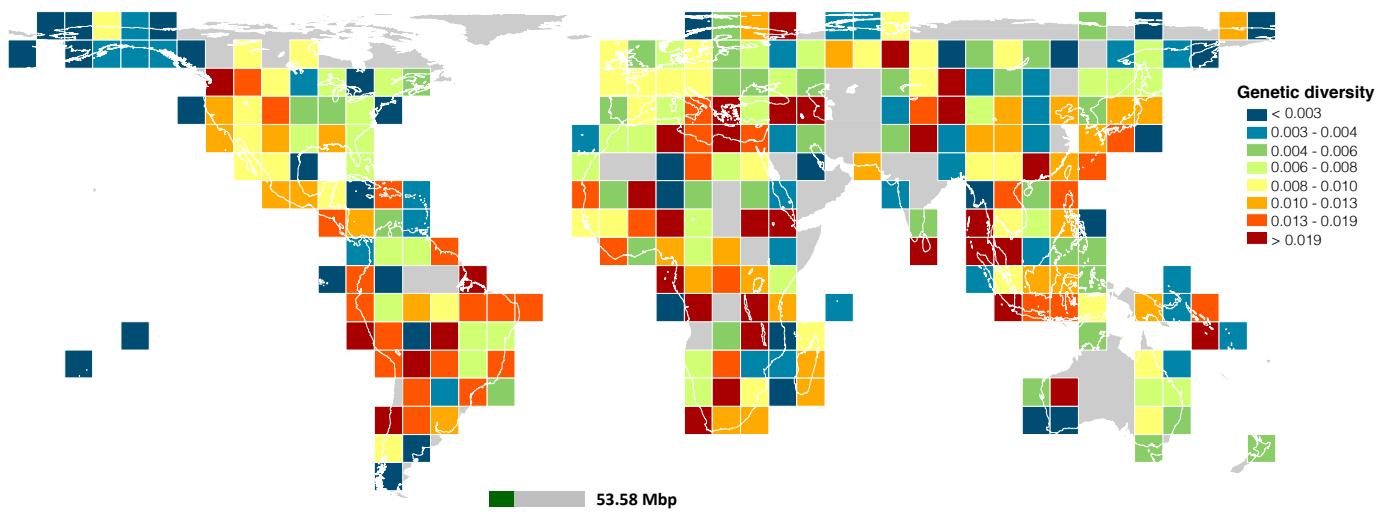
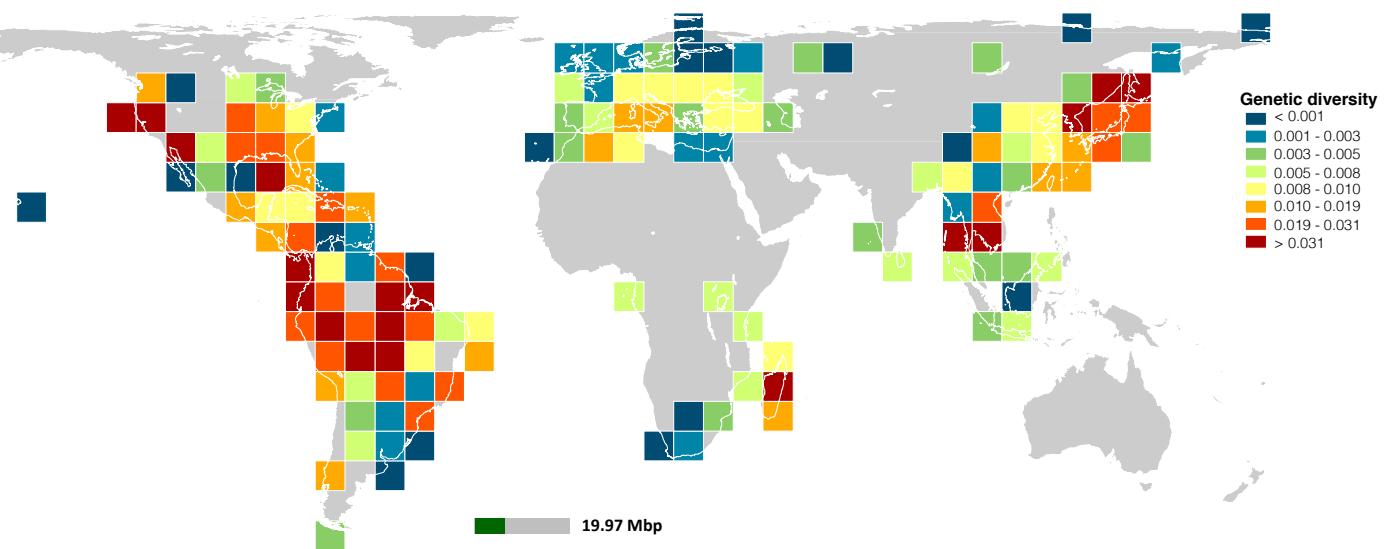
Upper percentiles



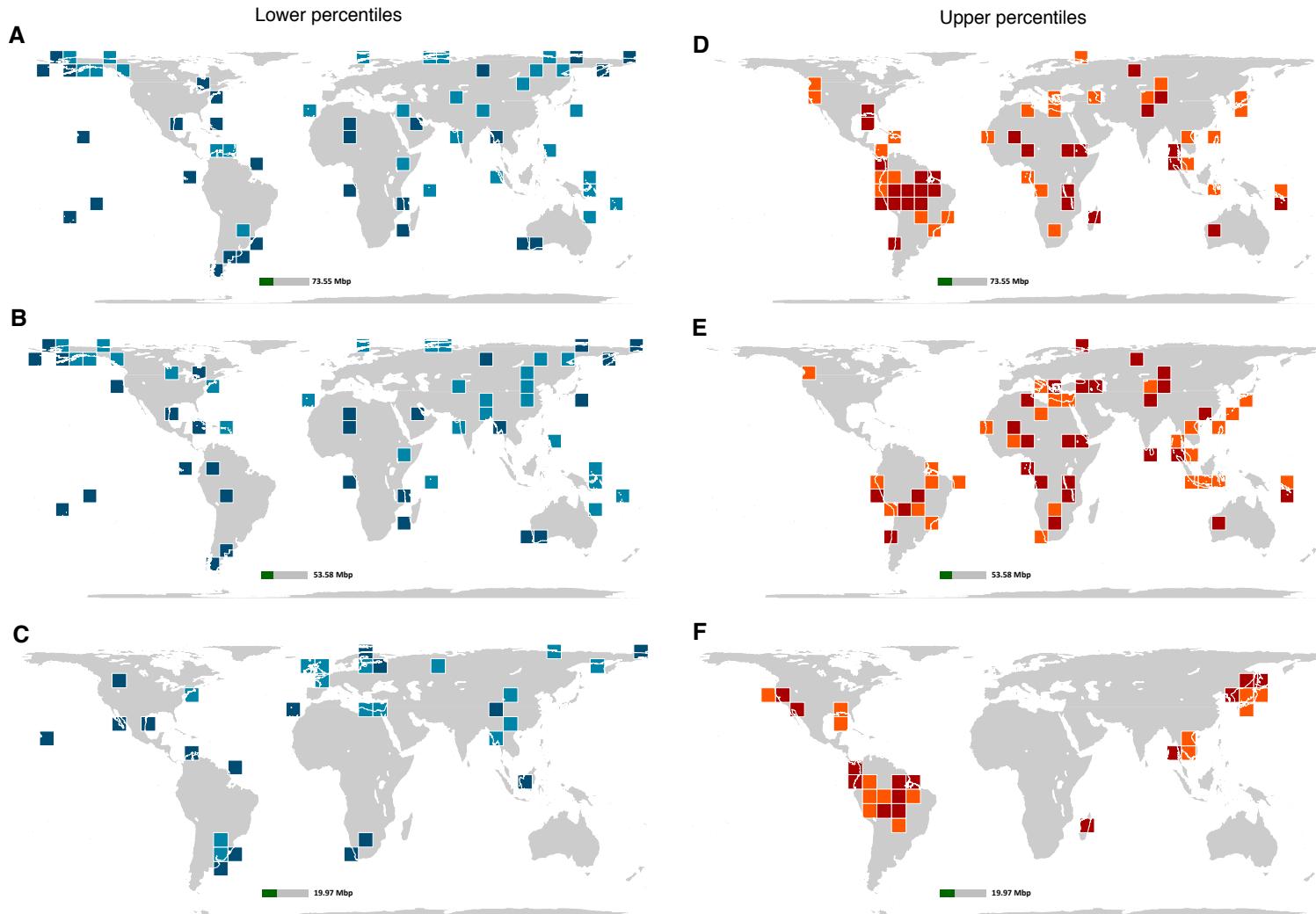
Lower percentiles

**B****E****C****F****Fig. S16 Global distribution of upper and lower percentiles of genetic diversity.**

Distribution of upper and lower percentiles of genetic diversity of combined taxon (**A** and **D**), terrestrial mammals (**B** and **E**) and amphibians (**C** and **F**). In the upper percentile maps (A-C), red represents the 90<sup>th</sup> percentile and orange the 80<sup>th</sup> percentile. In the lower percentile maps (D-F), dark blue represents the 10<sup>th</sup> percentile and light blue the 20<sup>th</sup> percentile. Percentiles for mammals and amphibians combined: 90<sup>th</sup> percentile = 0.023, 80<sup>th</sup> percentile = 0.015, 20<sup>th</sup> percentile = 0.003, 10<sup>th</sup> percentile = 0.001; mammals alone: 90<sup>th</sup> percentile = 0.019, 80<sup>th</sup> percentile = 0.013, 20<sup>th</sup> percentile = 0.003, 10<sup>th</sup> percentile = 0.001; amphibians alone: 90<sup>th</sup> percentile = 0.031, 80<sup>th</sup> percentile = 0.022, 20<sup>th</sup> percentile = 0.002, 10<sup>th</sup> percentile = 0.00.

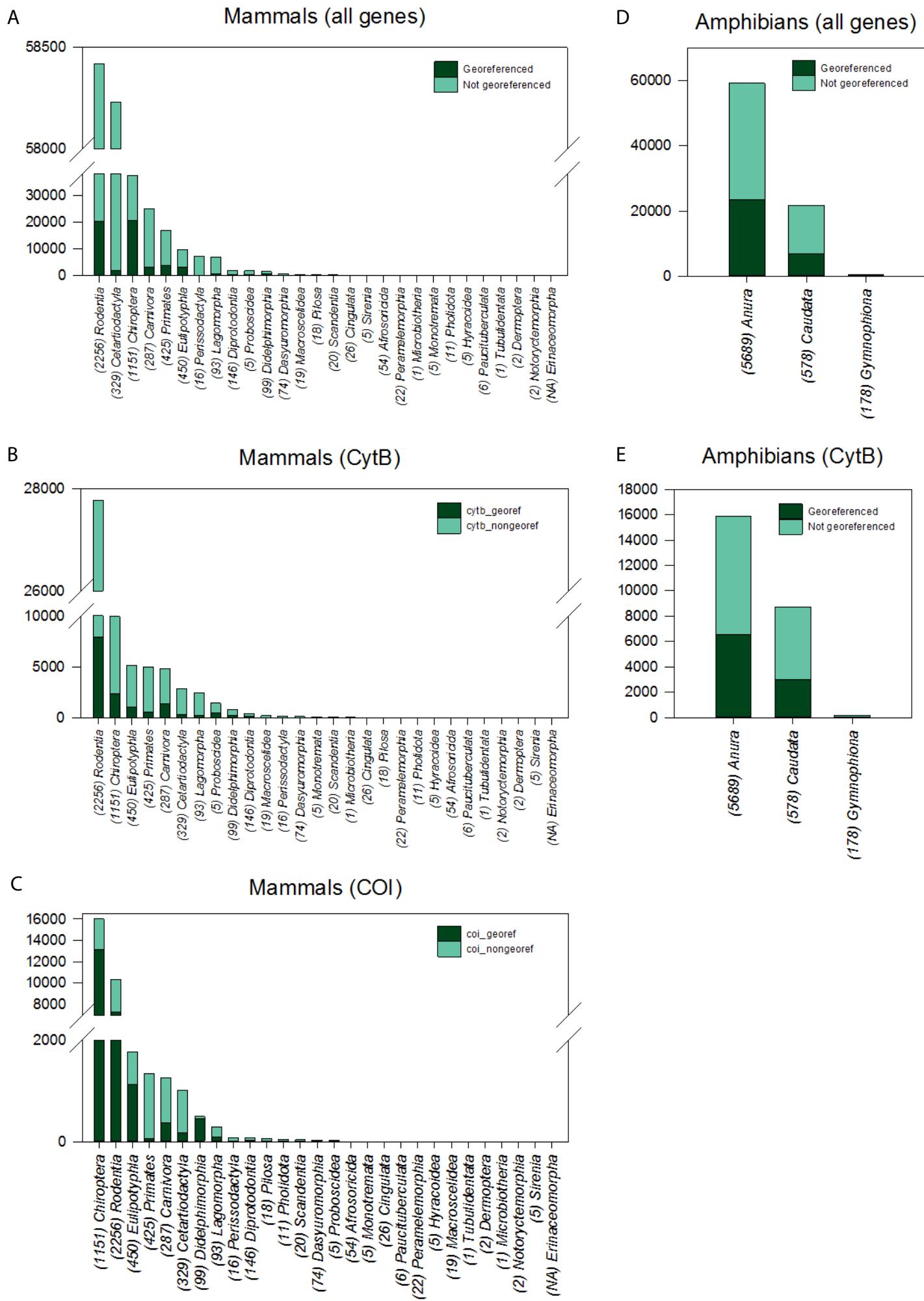
**A Combined taxon****B Mammals****C Amphibians**

**Fig. S17 Global distribution of genetic diversity as in Fig. 1 of the main text but at a coarser grain size (771.9 Km). Average number of mutations per base pair for cytochrome b across species of terrestrial mammals and amphibians together (A), terrestrial mammals alone (B) and amphibians alone (C).** Colors represent quantiles (8 quantiles). The gray bar below each map represents the total number of cytochrome b base pairs retrieved from GenBank and BOLD, and the green bar the number of georeferenced base pairs used to estimate the global distribution of genetic diversity.



**Fig. S18 Global distribution of upper and lower percentiles of genetic diversity as shown in Fig. S16, but using a coarser grain size (771.9 km).**

Distribution of upper and lower percentiles of genetic diversity of combined taxon (**A** and **D**), terrestrial mammals (**B** and **E**) and amphibians (**C** and **F**). In the upper percentile maps (D-F), red represents the 90<sup>th</sup> percentile and orange the 80<sup>th</sup> percentile. In the lower percentile maps (A-C), dark blue represents the 10<sup>th</sup> percentile and light blue the 20<sup>th</sup> percentile.



**Fig. S19 Number of mitochondrial sequences per taxonomic order for mammals (A-C) and amphibians (D-E), dowloaded from GenBank and IBOLD.**

## References

1. R. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974).
2. R. A. Fischer, *The Genetical Theory of Natural Selection* (Oxford Univ. Press, 1930).
3. R. Frankham, Inbreeding and extinction: A threshold effect. *Conserv. Biol.* **9**, 792–799 (1995). [doi:10.1046/j.1523-1739.1995.09040792.x](https://doi.org/10.1046/j.1523-1739.1995.09040792.x)
4. I. Saccheri, M. Kuussaari, M. Kankare, P. Vikman, W. Fortelius, I. Hanski, Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**, 491–494 (1998). [doi:10.1038/33136](https://doi.org/10.1038/33136)
5. D. Newman, D. Pilson, Increased probability of extinction due to decreased genetic effective population size: Experimental populations of *Clarkia pulchella*. *Evolution* **51**, 354–362 (1997). [doi:10.2307/2411107](https://doi.org/10.2307/2411107)
6. D. Spielman, B. W. Brook, R. Frankham, Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15261–15264 (2004). [Medline doi:10.1073/pnas.0403809101](https://doi.org/10.1073/pnas.0403809101)
7. M. Vellend, The consequences of genetic diversity in competitive communities. *Ecology* **87**, 304–311 (2006). [Medline doi:10.1890/05-0173](https://doi.org/10.1890/05-0173)
8. G. M. Crutsinger, M. D. Collins, J. A. Fordyce, Z. Gompert, C. C. Nice, N. J. Sanders, Plant genotypic diversity predicts community structure and governs an ecosystem process. *Science* **313**, 966–968 (2006). [Medline doi:10.1126/science.1128326](https://doi.org/10.1126/science.1128326)
9. J. K. Bailey, J. A. Schweitzer, F. Ubeda, J. Koricheva, C. J. LeRoy, M. D. Madritch, B. J. Rehill, R. K. Bangert, D. G. Fischer, G. J. Allan, T. G. Whitham, From genes to ecosystems: A synthesis of the effects of plant genetic factors across levels of organization. *Philos. Trans. R. Soc. London Ser. B* **364**, 1607–1616 (2009). [Medline doi:10.1098/rstb.2008.0336](https://doi.org/10.1098/rstb.2008.0336)
10. R. M. May, J. Godfrey, Biological diversity: Differences between land and sea. *Philos. Trans. R. Soc. London Ser. B* **343**, 105–111 (1994). [doi:10.1098/rstb.1994.0014](https://doi.org/10.1098/rstb.1994.0014)
11. M. C. Urban, Accelerating extinction risk from climate change. *Science* **348**, 571–573 (2015). [Medline doi:10.1126/science.aaa4984](https://doi.org/10.1126/science.aaa4984)
12. O. Hoegh-Guldberg, J. F. Bruno, The impact of climate change on the world's marine ecosystems. *Science* **328**, 1523–1528 (2010). [Medline doi:10.1126/science.1189930](https://doi.org/10.1126/science.1189930)
13. C. Moritz, R. Agudo, The future of species under climate change: Resilience or decline? *Science* **341**, 504–508 (2013). [Medline doi:10.1126/science.1237190](https://doi.org/10.1126/science.1237190)
14. R. Menéndez, A. G. Megías, J. K. Hill, B. Braschler, S. G. Willis, Y. Collingham, R. Fox, D. B. Roy, C. D. Thomas, Species richness changes lag behind climate change. *Proc. R. Soc. London Ser. B* **273**, 1465–1470 (2006). [Medline doi:10.1098/rspb.2006.3484](https://doi.org/10.1098/rspb.2006.3484)

15. H. M. Pereira, L. M. Navarro, I. S. Martins, Global biodiversity change: The bad, the good, and the unknown. *Annu. Rev. Environ. Resour.* **37**, 25–50 (2012). [doi:10.1146/annurev-environ-042911-093511](https://doi.org/10.1146/annurev-environ-042911-093511)
16. G. Hewitt, The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000). [Medline doi:10.1038/35016000](https://www.ncbi.nlm.nih.gov/pubmed/1091000)
17. W. Steffen, K. Richardson, J. Rockström, S. E. Cornell, I. Fetzer, E. M. Bennett, R. Biggs, S. R. Carpenter, W. de Vries, C. A. de Wit, C. Folke, D. Gerten, J. Heinke, G. M. Mace, L. M. Persson, V. Ramanathan, B. Reyers, S. Sörlin, Planetary boundaries: Guiding human development on a changing planet. *Science* **347**, 1259855 (2015). [10.1126/science.1259855](https://doi.org/10.1126/science.1259855) [Medline doi:10.1126/science.1259855](https://www.ncbi.nlm.nih.gov/pubmed/26251500)
18. H. M. Pereira, S. Ferrier, M. Walters, G. N. Geller, R. H. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. Butchart, A. C. Cardoso, N. C. Coops, E. Dulloo, D. P. Faith, J. Freyhof, R. D. Gregory, C. Heip, R. Höft, G. Hurt, W. Jetz, D. S. Karp, M. A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B. Reyers, R. Sayre, J. P. Scharlemann, S. N. Stuart, E. Turak, M. Walpole, M. Wegmann, Essential biodiversity variables. *Science* **339**, 277–278 (2013). [Medline doi:10.1126/science.1229931](https://www.ncbi.nlm.nih.gov/pubmed/23753111)
19. G. M. Hewitt, The structure of biodiversity—insights from molecular phylogeography. *Front. Zool.* **1**, 4 (2004). [Medline doi:10.1186/1742-9994-1-4](https://www.ncbi.nlm.nih.gov/pubmed/15244144)
20. L. B. Beheregaray, Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Mol. Ecol.* **17**, 3754–3774 (2008). [Medline](https://www.ncbi.nlm.nih.gov/pubmed/18370000)
21. See supplementary materials on *Science* Online.
22. D. R. Vieites, M.-S. Min, D. B. Wake, Rapid diversification and dispersal during periods of global warming by plethodontid salamanders. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19903–19907 (2007). [Medline doi:10.1073/pnas.0705056104](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1936030/)
23. R. Grenyer, C. D. Orme, S. F. Jackson, G. H. Thomas, R. G. Davies, T. J. Davies, K. E. Jones, V. A. Olson, R. S. Ridgely, P. C. Rasmussen, T. S. Ding, P. M. Bennett, T. M. Blackburn, K. J. Gaston, J. L. Gittleman, I. P. Owens, Global distribution and conservation of rare and threatened vertebrates. *Nature* **444**, 93–96 (2006). [Medline doi:10.1038/nature05237](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470000/)
24. K. Rohde, Latitudinal gradients in species diversity: The search for the primary cause. *Oikos* **65**, 514–527 (1992). [doi:10.2307/3545569](https://doi.org/10.2307/3545569)
25. G. G. Mittelbach, D. W. Schemske, H. V. Cornell, A. P. Allen, J. M. Brown, M. B. Bush, S. P. Harrison, A. H. Hurlbert, N. Knowlton, H. A. Lessios, C. M. McCain, A. R. McCune, L. A. McDade, M. A. McPeek, T. J. Near, T. D. Price, R. E. Ricklefs, K. Roy, D. F. Sax, D. Schlüter, J. M. Sobel, M. Turelli, Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecol. Lett.* **10**, 315–331 (2007). [Medline doi:10.1111/j.1461-0248.2007.01020.x](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1936030/)

26. P. V. A. Fine, Ecological and evolutionary drivers of geographic variation in species diversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 369–392 (2015). [doi:10.1146/annurev-ecolsys-112414-054102](https://doi.org/10.1146/annurev-ecolsys-112414-054102)
27. R. Dirzo, H. S. Young, M. Galetti, G. Ceballos, N. J. Isaac, B. Collen, Defaunation in the Anthropocene. *Science* **345**, 401–406 (2014). [Medline](#) [doi:10.1126/science.1251817](https://doi.org/10.1126/science.1251817)
28. C. N. Waters, J. Zalasiewicz, C. Summerhayes, A. D. Barnosky, C. Poirier, A. Gałuszka, A. Cebarreta, M. Edgeworth, E. C. Ellis, M. Ellis, C. Jeandel, R. Leinfelder, J. R. McNeill, D. Richter, W. Steffen, J. Syvitski, D. Vidas, M. Wagreich, M. Williams, A. Zhisheng, J. Grinevald, E. Odada, N. Oreskes, A. P. Wolfe, The Anthropocene is functionally and stratigraphically distinct from the Holocene. *Science* **351**, aad2622 (2016). [10.1126/science.aad2622](https://doi.org/10.1126/science.aad2622) [Medline](#) [doi:10.1126/science.aad2622](https://doi.org/10.1126/science.aad2622)
29. E. C. Ellis, K. K. Goldewijk, S. Siebert, D. Lightman, N. Ramankutty, Anthropogenic transformation of the biomes, 1700 to 2000. *Glob. Ecol. Biogeogr.* **19**, 589–606 (2010). [doi:10.1111/j.1466-8238.2010.00540.x](https://doi.org/10.1111/j.1466-8238.2010.00540.x)
30. Data available at <http://macroecology.ku.dk/resources/imapgenes>.
31. M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012). [Medline](#) [doi:10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199)
32. J. V. Lopez, N. Yuhki, R. Masuda, W. Modi, S. J. O'Brien, Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* **39**, 174–190 (1994). [Medline](#)
33. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). [Medline](#) [doi:10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340)
34. J. Bezanson, A. Edelman, S. Karpinski, V. B. Shah, <https://arxiv.org/abs/1411.1607> (2015).
35. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2015); [www.r-project.org](http://www.r-project.org).
36. M. Nei, *Molecular Evolutionary Genetics* (Arizona State University, 1987).
37. M. Nei, W. H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5269–5273 (1979). [Medline](#) [doi:10.1073/pnas.76.10.5269](https://doi.org/10.1073/pnas.76.10.5269)
38. F. Tajima, Statistical analysis of DNA polymorphism. *Jpn. J. Genet.* **68**, 567–595 (1993). [Medline](#) [doi:10.1266/jgg.68.567](https://doi.org/10.1266/jgg.68.567)
39. QGIS Geographic Information System, Open Source Geospatial Foundation Project (2011); <http://qgis.osgeo.org>.

40. D. M. Olson, E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. N. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand, J. C. Morrison, C. J. Loucks, T. F. Allnutt, T. H. Ricketts, Y. Kura, J. F. Lamoreux, W. W. Wettengele, P. Hedao, K. R. Kassem, Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* **51**, 933–938 (2001). [doi:10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
41. IUCN Red List of Threatened Species (amphibians), version 2014.4; [www.iucnredlist.org/amphibians](http://www.iucnredlist.org/amphibians).
42. IUCN Red List of Threatened Species (mammals), version 2014.4; [www.iucnredlist.org/mammals](http://www.iucnredlist.org/mammals).
43. D. R. Frost, Amphibian Species of the World: An Online Reference, version 6.0 (American Museum of Natural History, New York); <http://research.amnh.org/herpetology/amphibia/index.html>.
44. D. E. Wilson, D. M. Reeder, Eds., *Mammal Species of the World: A Taxonomic and Geographic Reference* (Johns Hopkins Univ. Press, ed. 3, 2005).
45. M. K. Borregaard, C. Rahbek, J. Fjeldså, J. L. Parra, R. J. Whittaker, C. H. Graham, Node-based analysis of species distributions. *Methods Ecol. Evol.* **5**, 1225–1235 (2014). [doi:10.1111/2041-210X.12283](https://doi.org/10.1111/2041-210X.12283)
46. P. Dutilleul, P. Clifford, S. Richardson, D. Hemon, Modifying the *t* test for assessing the correlation between two spatial processes. *Biometrics* **49**, 305 (1993). [doi:10.2307/2532625](https://doi.org/10.2307/2532625)
47. F. Osorio, R. Vallejos, SpatialPack: Package for Analysis of Spatial Data, R Package version 0.2-3 (2014); <http://cran.r-project.org/package=SpatialPack>.
48. S. L. P. Ferrari, F. Cribari-Neto, Beta Regression for modelling rates and proportions. *J. Appl. Stat.* **31**, 799–815 (2004). [doi:10.1080/0266476042000214501](https://doi.org/10.1080/0266476042000214501)
49. M. Smithson, J. Verkuilen, A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol. Methods* **11**, 54–71 (2006). [Medline doi:10.1037/1082-989X.11.1.54](https://doi.org/10.1037/1082-989X.11.1.54)
50. F. Cribari-Neto, A. Zeileis, Beta Regression in R. *J. Stat. Softw.* **34**, 1–24 (2010). [doi:10.18637/jss.v034.i02](https://doi.org/10.18637/jss.v034.i02)
51. V. E. Seshan, clinfun: Clinical Trial Design and Data Analysis Functions, R Package version 1.0.11 (2015); <http://CRAN.R-project.org/package=clinfun>.
52. R. Mac Nally, Regression and model-building in conservation biology, biogeography and ecology: The distinction between—and reconciliation of—“predictive” and “explanatory” models. *Biodivers. Conserv.* **9**, 655–671 (2000). [doi:10.1023/A:1008985925162](https://doi.org/10.1023/A:1008985925162)

53. C. Walch, R. Mac Nally, *hier.part*: Hierarchical Partitioning, R Package version 1.0-4 (2013);  
<http://cran.r-project.org/package=hier.part>.