BIODIVERSITY

An Anthropocene map of genetic diversity

Andreia Miraldo,[1]* Sen Li,[1] Michael K. Borregaard,[1] Alexander Flórez-Rodríguez,[1]
Shyam Gopalakrishnan,[2] Mirnesa Rizvanovic,[1] Zhiheng Wang,[1,3] Carsten Rahbek,[1,4]
Katharine A. Marske,[1] David Nogués-Bravo[1]*

The Anthropocene is witnessing a loss of biodiversity, with well-documented declines in the diversity of ecosystems and species. For intraspecific genetic diversity, however, we lack even basic knowledge on its global distribution. We georeferenced 92,801 mitochondrial sequences for >4500 species of terrestrial mammals and amphibians, and found that genetic diversity is 27% higher in the tropics than in nontropical regions. Overall, habitats that are more affected by humans hold less genetic diversity than wilder regions, although results for mammals are sensitive to choice of genetic locus. Our study associates geographic coordinates with publicly available genetic sequences at a massive scale, yielding an opportunity to investigate both the drivers of this component of biodiversity and the genetic consequences of the anthropogenic modification of nature.

I ntraspecific genetic diversity, or the amount of genetic variation among individuals within a species, provides the critical basis for evolutionary change (*1*, *2*), such as adaptation to new environmental conditions. Because it both reflects and influences processes at the population (*3*, *4*), species (*5*, *6*), community (*7*), and ecosystem levels (*8*, *9*), genetic diversity is often considered the most fundamental dimension of biodiversity (*10*). Although the global distribution of species and ecosystems and their responses to global change have been extensively investigated in recent decades (*11–14*), the global distribution of intraspecific genetic diversity is still largely unknown (*15*). Genetic diversity has already diminished for many species as a result of abrupt climatic fluctuations and human activity across the Late Quaternary (*16*), and this trend may be accelerating today. Knowledge of the global distribution of genetic diversity is therefore of critical importance if we want to fully understand the long-term impacts of human-induced global changes in climate and land use on genetic diversity, evaluate the potential of species to adapt to global change in the Anthropocene, and ultimately to succeed in halting biodiversity loss. Indeed, genetic diversity has recently been identified as an essential biodiversity variable (*17*) required to monitor biosphere integrity, one of the nine environmental planetary boundaries within which humanity can safely operate (*18*).

[1]Center for Macroecology, Evolution, and Climate, Natural History Museum of Denmark, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark. [2]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark. [3]Department of Ecology and Key Laboratory for Earth Surface Processes of the Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. [4]Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.
*Corresponding author. Email: andreia.miraldo@snm.ku.dk (A.M.); dnogues@snm.ku.dk (D.N.-B.)

Most of the current knowledge about the spatial distribution of intraspecific genetic diversity comes from the wealth of phylogeographic studies accumulated over the past 25 years. Although these studies have reported some emergent spatial patterns of genetic diversity at regional scales (*16*, *19*), few regions have been studied as intensively as temperate Europe and North America (*20*), and truly global patterns thus remain to be identified. However, millions of genetic sequences from a diverse array of studies have been deposited in public repositories during this same period (>180 million in GenBank alone), constituting an exceptional source of primary genetic data that could be used to explore global geographical patterns. While impressive in number, the majority of these sequences are not accompanied by geographic coordinates (>85% of all sequences in GenBank), and their integration with other biodiversity information would require manually reviewing the thousands of individual papers in which they are published.

Here, we took advantage of sequences publicly available in the GenBank and BOLD repositories and devised bioinformatic tools to map the global distribution of genetic diversity (*21*). We have attached geographic coordinates to a total of 92,801 mitochondrial sequences (31,029 for amphibians and 61,772 for terrestrial mammals), representing 38% and 27% of available sequences for amphibians and mammals, respectively. Using 86,406 *cytochrome b* (*cytb*) sequences, we performed species-specific sequence alignments for 4675 species and calculated nucleotide diversity per site for each species through pairwise comparisons of georeferenced aligned sequences (24,479 *cytb* sequences from 1992 species). To map the average number of genetic mutations globally (Fig. 1), we estimated the genetic diversity of each equal area grid cell (~150,000 km²) by averaging nucleotide diversity per site across all species present (*21*).

The global map of genetic diversity for mammals and amphibians together (Fig. 1A) shows that the tropical Andes and Amazonia harbor some of the highest levels of genetic diversity (90th percentile: >0.023; fig. S16). Other regions with high genetic diversity include the subtropical parts of South Africa for mammals (80th percentile: >0.013) and eastern parts of the Sino-Japanese region for amphibians (80th percentile: >0.022; fig. S16). Within the temperate regions of the planet, western North America also contains high levels of genetic diversity, coinciding with the high mammalian species richness there (*22*), whereas eastern North America harbors high levels of genetic diversity in a region that is the global center of speciation and species richness for salamanders (*23*). These patterns are robust to spatial variation in sampling intensity, as demonstrated by rarefaction analyses (*21*), but the high variability of genetic diversity across grid cells with low sample size requires a cautious interpretation of local patterns. The same analysis based on an alternative gene for mammals, *cytochrome oxidase subunit I* (22,762 sequences across 966 species), shows that although genetic diversity values between loci are not correlated for individual grid cells ($r = 0.094$, $P = 0.151$; fig. S1), they are highly congruent across latitudinal bands ($r = 0.79$, $P = 0.001$; fig. S4).
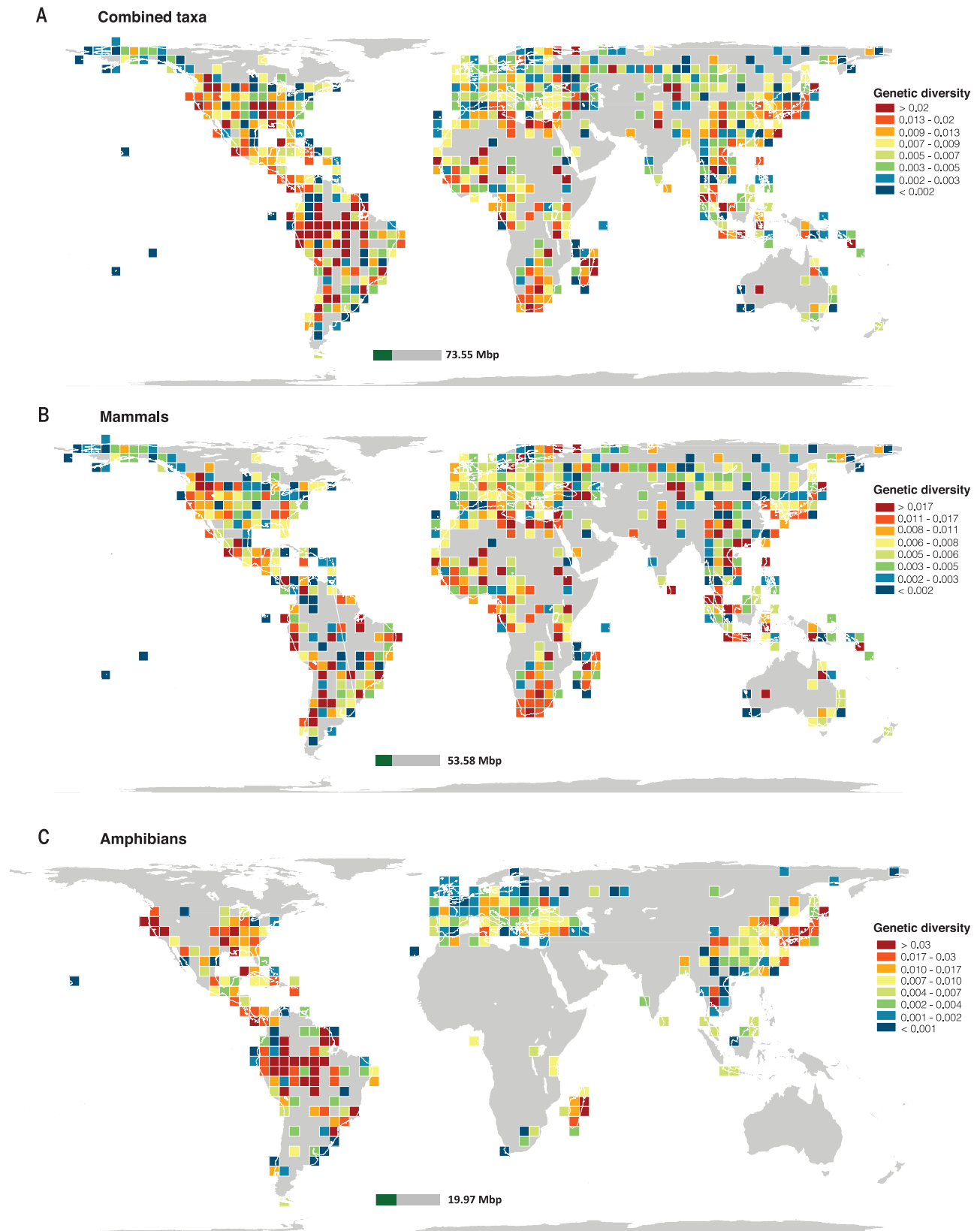
Sequence availability and taxonomic coverage vary greatly across the globe. The maps of ignorance (Fig. 2 and fig. S10) illustrate the spatial distribution of key gaps in both types of coverage. Unsurprisingly, the majority of knowledge comes from western Europe, North America, and far East Asia, plus individual regions that have recently been the focus of much biogeographic and phylogeographic research, such as Madagascar. Interestingly, western Europe supports some of the lowest levels of genetic diversity (20th percentile: <0.0015) for amphibians (Fig. 1C and fig. S16), although this is one of the best-sampled regions in terms of both sequence availability and taxonomic coverage (Fig. 2C and fig. S10C). This result suggests that our estimates of genetic diversity are robust to the observed geographical bias in data availability, a conclusion that is corroborated by our sensitivity analyses (figs. S7 and S8).

Despite the lower density of data in tropical regions relative to temperate regions of the Northern Hemisphere, we nonetheless identify a pattern of higher genetic diversity in the tropics with a decrease toward the poles, which broadly mirrors the latitudinal gradient of species richness (mammals: pseudo-$r^2 = 0.80$, quadratic $P = 0.001$; amphibians: pseudo-$r^2 = 0.73$, quadratic $P = 0.001$) (Fig. 3, A and C). It has long been speculated that higher temperatures in the tropics may result in increased evolutionary rates and that this may be the driver for the latitudinal species richness gradient (*24*). Our results are in accordance with the expectation of the proposed mechanism of the evolutionary speed hypothesis, which links biodiversity to temperature through its direct effect on mutation rates and generation time, which in turn may increase the rates of population genetic divergence and speciation (*24*). Mittelbach *et al.*

A Combined taxa

B Mammals

C Amphibians

Fig. 1. Global distribution of genetic diversity. (**A** to **C**) Average number of mutations per base pair for *cytb* across species of terrestrial mammals and amphibians together (A), terrestrial mammals alone (B), and amphibians alone (C). Different colors represent eight quantiles. The gray bar below each map represents the total number of *cytb* base pairs retrieved from GenBank and BOLD; the green bar shows the number of georeferenced base pairs used to estimate the global distribution of genetic diversity.

Fig. 2. Global distribution of knowledge and ignorance. (**A** to **C**) Distribution using all mitochondrial 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 mitochondrial sequence is available for that cell. The *y* axis indicates sequence availability [i.e., the total number of georeferenced mitochondrial base pairs (log scale) within each grid cell]. In each map, violet colors indicate regions of low taxonomic coverage and relatively few geo referenced base pairs; green colors indicate higher taxonomic coverage and higher numbers of base pairs; light blue colors indicate high taxonomic coverage and few base pairs; red colors indicate low taxonomic coverage and high numbers of base pairs.

A Combined taxa

B Mammals

C Amphibians
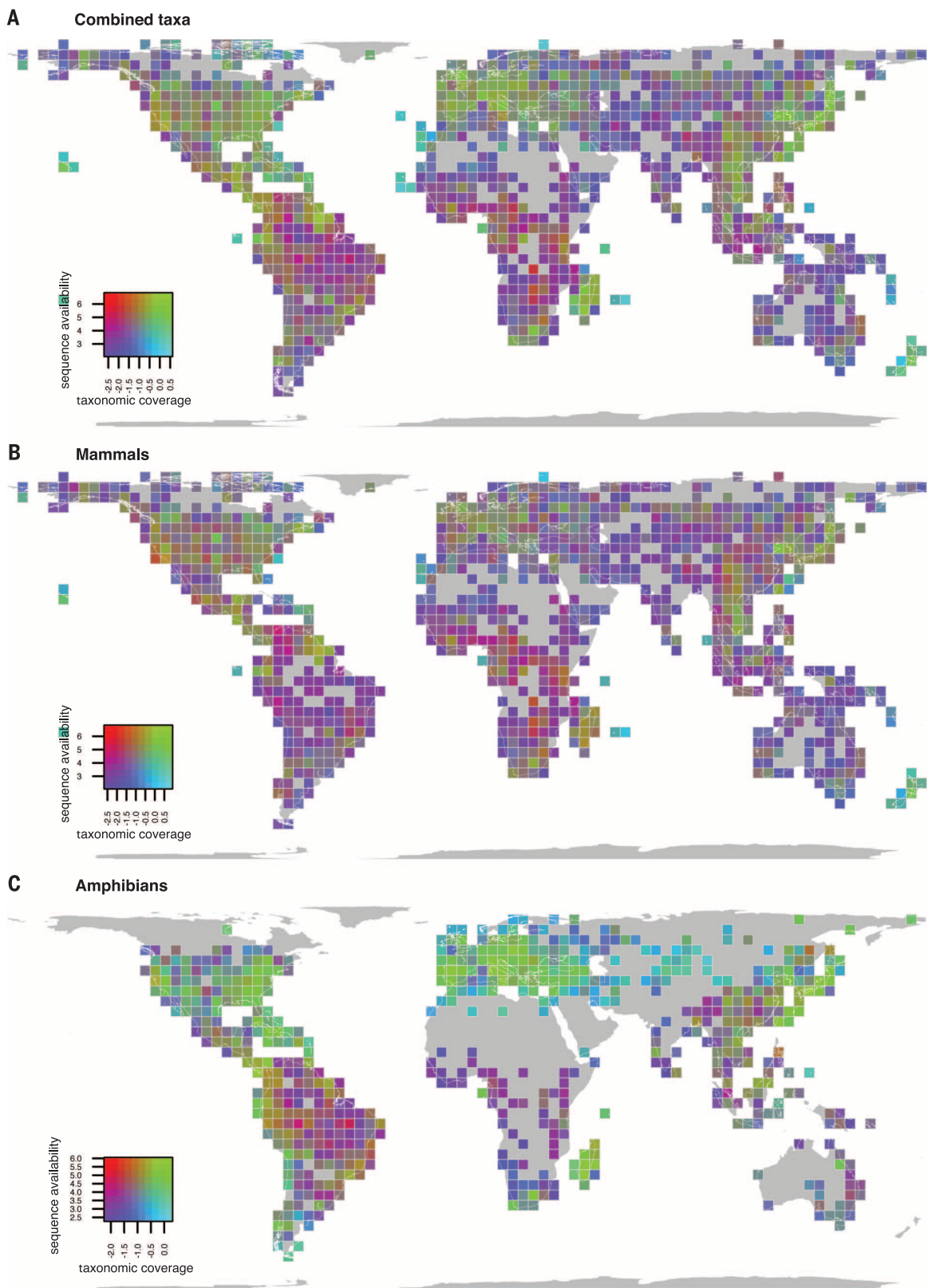
(*25*) observed that "several formidable challenges to the evolutionary speed hypothesis remain," including the lack of appropriate data at a scale from which generalities may be inferred and tested. The data presented here open the door to testing this and other potential mechanisms driving biological diversity in the tropics, such as climatic stability or the complex relationship among geographical area, water-energy dynamics, and evolutionary time (*26*).

There is a growing consensus that humans have transformed habitats and have affected ecosystems
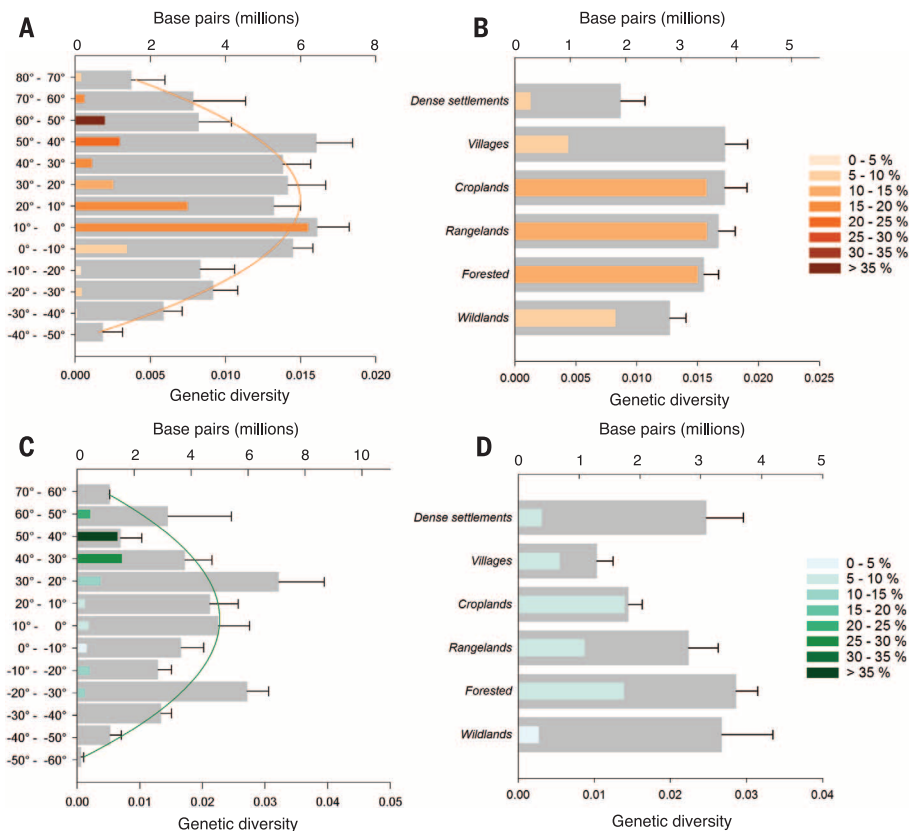
Fig. 3. Distribution of genetic diversity across latitudes and anthromes for mitochondrial *cytb*. (**A** to **D**) Gray bars (lower *x* axis) indicate the average number of mutations per base pair for *cytb* (means ± SD) across species of terrestrial mammals [(A) and (B)] and amphibians [(C) and (D)] in each latitudinal band [(A) and (C)] and anthrome [(B) and (D)]. The length of colored bars (upper *x* axis) indicates the number of *cytb* base pairs used in the calculation of genetic diversity (GD); the depth of color shading denotes the percentage of species used in the calculation of GD relative to species richness of each category. Quadratic polynomial models indicate a peak of GD in the tropics (mammals, pseudo-r^2 = 0.80; amphibians, pseudo-r^2 = 0.73). Error bars represent SD from the mean of the bootstrapping analysis.

across most of the terrestrial biosphere (*27*, *28*). This transformation has been recently mapped on the basis of human population density and land use, leading to a classification of the terrestrial world into anthropogenic ecosystems, or anthromes (*29*). This classification allows exploration of potential anthropogenic impacts on the distribution of genetic diversity, independently for amphibians and mammals. For amphibians, we find an increase in genetic diversity as we move from anthromes more heavily transformed by humans—dense settlements, villages, and croplands—toward those less affected by human pressure, including rangelands, forested ecosystems and wildlands (Jonckheere-Terpstra test, JT statistic = 55,006.5, *P* = 0.004, increasing trend; Fig. 3D and S14A). For mammals, we detect the same pattern of increasing genetic diversity toward anthromes less affected by human pressure for the *co1* gene (Jonckheere-Terpstra test, JT statistic = 352,761, *P* = 0.025, increasing trend; figs. S4B and S14B); however, differences in genetic diversity across anthromes are not statistically significant for

the *cytb* gene (Kruskal-Wallis test, χ^2 = 9.696, df = 5, *P* = 0.084; Fig. 3B and fig. S14C).

This study does not intend to explain the processes responsible for the higher levels of genetic diversity in the tropics, nor its relationship with human impacts in the Anthropocene. However, it reveals the very limited knowledge of the global distribution of biological diversity at its most fundamental dimension, the genetic diversity level. As Fig. 2 shows, the regions of least knowledge are also those that harbor some of the highest numbers of species on Earth. Consequently, there is a need to develop data-mining algorithms to georeference the millions of sequences already available in public repositories, as well as better strategies to curate future phylogeographic data. Moreover, it is important to hasten the collection of new genetic data in the field to cover undersampled regions. To encourage such efforts, we make the coordinates of the georeferenced sequences—representing 30% of the sequences available in public repositories for these taxa and genetic loci—freely available at the iMapGenes website (*30*), along with the analytical tools to estimate spatial patterns of intraspecific diversity.

REFERENCES AND NOTES

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, *Conserv. Biol.* **9**, 792–799 (1995).
4. I. Saccheri *et al.*, *Nature* **392**, 491–494 (1998).
5. D. Newman, D. Pilson, *Evolution* **51**, 354–362 (1997).
6. D. Spielman, B. W. Brook, R. Frankham, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15261–15264 (2004).
7. M. Vellend, *Ecology* **87**, 304–311 (2006).
8. G. M. Crutsinger *et al.*, *Science* **313**, 966–968 (2006).
9. J. K. Bailey *et al.*, *Philos. Trans. R. Soc. London Ser. B* **364**, 1607–1616 (2009).
10. R. M. May, J. Godfrey, *Philos. Trans. R. Soc. London Ser. B* **343**, 105–111 (1994).
11. M. C. Urban, *Science* **348**, 571–573 (2015).
12. O. Hoegh-Guldberg, J. F. Bruno, *Science* **328**, 1523–1528 (2010).
13. C. Moritz, R. Agudo, *Science* **341**, 504–508 (2013).
14. R. Menéndez *et al.*, *Proc. R. Soc. London Ser. B* **273**, 1465–1470 (2006).
15. H. M. Pereira, L. M. Navarro, I. S. Martins, *Annu. Rev. Environ. Resour.* **37**, 25–50 (2012).
16. G. Hewitt, *Nature* **405**, 907–913 (2000).
17. W. Steffen *et al.*, *Science* **347**, 1259855 (2015).
18. H. M. Pereira *et al.*, *Science* **339**, 277–278 (2013).
19. G. M. Hewitt, *Front. Zool.* **1**, 4 (2004).
20. L. B. Beheregaray, *Mol. Ecol.* **17**, 3754–3774 (2008).
21. See supplementary materials on *Science* Online.
22. D. R. Vieites, M.-S. Min, D. B. Wake, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19903–19907 (2007).
23. R. Grenyer *et al.*, *Nature* **444**, 93–96 (2006).
24. K. Rohde, *Oikos* **65**, 514–527 (1992).
25. G. G. Mittelbach *et al.*, *Ecol. Lett.* **10**, 315–331 (2007).
26. P. V. A. Fine, *Annu. Rev. Ecol. Evol. Syst.* **46**, 369–392 (2015).
27. R. Dirzo *et al.*, *Science* **345**, 401–406 (2014).
28. C. N. Waters *et al.*, *Science* **351**, aad2622 (2016).
29. E. C. Ellis, K. K. Goldewijk, S. Siebert, D. Lightman, N. Ramankutty, *Glob. Ecol. Biogeogr.* **19**, 589–606 (2010).
30. Data available at http://macroecology.ku.dk/resources/imapgenes.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/353/6307/1532/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S5
References (*31–53*)
Data files S1 and S2

Science

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 and David Nogués-Bravo

ARTICLE TOOLS	http://science.sciencemag.org/content/353/6307/1532
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2016/09/28/353.6307.1532.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/353/6307/1494.full
REFERENCES	This article cites 38 articles, 11 of which you can access for free http://science.sciencemag.org/content/353/6307/1532#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions