Uneven genetic data limits biodiversity assessments in protected areas globally

Authors: Ivan Paz-Vinas[1,2]*†, Amy G. Vandergast[3]*†, Chloé Schmidt[4], Deborah M Leigh[5], Simon Blanchet[6], René D. Clark[7], Eric D. Crandall[8], Hanne De Kort[9], Jeff Falgout[10], Colin J. Garroway[11], Eleana Karachaliou[11], Francine Kershaw[12], David O'Brien[13], Malin L. Pinsky[14,15], Gernot Segelbacher[16], Margaret E. Hunter[17]

Affiliations:

[1] Department of Biology, Colorado State University; 1878 Campus Delivery, Fort Collins, CO, 80523-1878, USA

[2] Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR5023 LEHNA; F-69622, Villeurbanne, France

[3] U.S. Geological Survey, Western Ecological Research Center; 4165 Spruance Road, Suite 200, San Diego, CA, 92101, USA

[4] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig; Puschstrasse 4, Leipzig, 04103, Germany

[5] Swiss Federal Research Institute for Forest, Snow, and Landscape Research WSL; Birmensdorf, Switzerland

[6] CNRS, Station d'Ecologie Théorique et Expérimentale; UAR 2029, 2 route du CNRS 09200 Moulis

[7] Department of Biology, Drexel University; 3245 Chestnut St. Philadelphia, PA, 19104, USA

[8] Department of Biology, Pennsylvania State University; 208 Mueller Laboratory University Park, PA 16802, USA

[9] Division of Ecology, Evolution and Biodiversity Conservation, Biology Department KU Leuven; 3001 Leuven, Belgium

[10] U.S. Geological Survey, Science Analytics and Synthesis; Denver, CO 80225, USA

[11] Department of Biological Sciences, University of Manitoba; 66 Chancellors Cir, Winnipeg, MB, R3T 2N2, Canada

[12] Oceans Division, Natural Resources Defense Council; 40 West 20th Street, New York, 10011, USA

[13] NatureScot, Great Glen House; Leachkin Road, Inverness, IV3 8NW, UK

[14] Department of Ecology and Evolutionary Biology, University of California Santa Cruz; Santa Cruz, CA, USA

[15] Department of Ecology, Evolution, and Natural Resources, Rutgers University; New Brunswick, NJ, USA

[16] Wildlife Ecology and Management; University Freiburg, Germany

[17] U.S. Geological Survey, Wetland and Aquatic Research Center; 7920 NW 71st Street, Gainesville, Florida 32653

* Corresponding authors: ivan.paz-vinas@univ-lyon1.fr and avandergast@usgs.gov
† These authors contributed equally to this work

Abstract: Increasing the extent of protected areas (PA) through 30x30 and other area-based conservation initiatives can help to achieve global biodiversity conservation goals across all biodiversity levels. However, intraspecific genetic variation, the foundational level of biodiversity, is rarely explicitly considered in PA design or quality performance assessments. Repurposing existing genetic data could rapidly inform area-based conservation planning and improve the preservation of genetic variation. Through a global compilation of population-level nuclear genetic data (>2 million individuals; 36,356 populations; 2,809 species), we identified both data-rich areas, and substantial geographic and taxonomic gaps. These gaps are within many protected areas and hotspots of species biodiversity, and may preclude robust protection of genetic diversity. Addressing data unevenness through efforts to collect, gather, harmonize and share genetic data globally could help support integration of genetic information into PA design, PA performance assessments, and genetically-oriented global conservation policies.

One-Sentence summary: Increased efforts to gather, harmonize and share existing population-level nuclear genetic data globally can help inform area-based conservation initiatives**.**

Main text:

Genetic variation within populations (hereafter, *genetic variation*) underlies population health, adaptive potential, species persistence, and ecosystem resilience (*1*). Yet, it has been predicted that 5.4%-6.5% (*2*) to >10% (*3*) of genetic variation, on average, may have been lost from species globally since the industrial revolution. Despite being the most fundamental level of biodiversity, genetic variation has historically been neglected by many global conservation initiatives (*1*) and excluded from virtually all protected area design and effectiveness assessments (*4*, *5*, but see *6*). The recently-adopted Kunming-Montreal Global Biodiversity Framework (KMGBF; *7*) is the first to set explicit goals and targets for maintaining and restoring genetic variation in wild species and for safeguarding their adaptive potential (KMGBF Goal A and Target 4; *7*). It also includes a target to increase the area protected globally to 30% by 2030 (i.e., KMGBF Target 3; *7*). Determining the baseline status of genetic variation in protected areas and beyond will be fundamental in measuring progress towards KMGBF genetic variation protection commitments and supporting conservation initiatives across the globe. A primary step towards informing genetic monitoring in protected areas is assessing the global availability and gaps of population-level nuclear genetic information. Previous global maps of genetic diversity (e.g., *8*) used spatial aggregations of mitochondrial DNA sequences, which are less sensitive to the fine-scale short-term population processes important for monitoring than population-level nuclear genetic data (*9* and references therein). Importantly, global maps of nuclear genetic data could also leverage available population-level genetic information to synergistically inform both global genetic conservation commitments (*1*, *7*) and other flagship area-based conservation efforts like the "30x30" initiative (e.g., KMGBF Target 3; *7*).

Protected areas (PAs, including marine protected areas and areas under other effective area-based conservation measures) are a cornerstone of biodiversity conservation efforts (*10*). Specifically, by preserving or connecting plant and animal populations, PAs can help safeguard genetic variation (*4*). Ideally, direct assessments of the representation of genetic

variation in PAs would rely on empirical genetic data collected across many populations and species from within and outside PAs (*6*). The exponential accumulation of genetic data in the last decades now enables repurposing intraspecific nuclear genetic data (*9*), which could rapidly inform PA design and effectiveness. However, the spatial and taxonomic distribution of available population-level nuclear genetic data remains unknown, as there is no singular global database for nuclear genetic variation at the population level that covers most major taxonomic groups.

We assessed the global availability of nuclear genetic data within and outside ~50.5 million km² of PAs, across all available biome types, by gathering geographical coordinates for nearly 2 million genotyped individuals forming 36,356 georeferenced groups of individuals (hereafter, *local populations*; see "Materials and Methods" section in the Supplementary Materials) from 2,809 species (Fig. 1; Table S1) and comparing these with the global database of protected areas. We present area assessments based on total area protected (km^2) rather than the number of PAs, because there is significant geographical overlap among individual PAs in the global protected areas database. Overall, 52% of the total area protected globally (64% terrestrial, 41% marine; Fig. 1) did not have at least one sampled local population (mean number of local populations *per* PA with genetic data = 2.57, median = 1; max. = 278). Some of the largest PAs had no, or limited sampling (e.g., Rapa Nui Island, Chile; Kermadec Islands, New Zealand; Marae Moana, Cook Islands; French Southern and Antarctic Lands National Nature Reserve; Fig. 1; Supplementary data 1). Furthermore, genetic data are currently lacking from important biodiversity hotspots like the Horn of Africa and the East Melanesian islands, highlighting the need to produce and gather genetic data within these areas (Fig. 1; Table S3).

In PAs where data exist, the density of local populations was generally low, with numerous geographical and taxonomic gaps (Figs. 1-2) that could limit range-wide assessments of genetic diversity and connectivity within and between PAs. Effectively evaluating genetic conservation would ideally include data from multiple species for multi-taxa planning efforts. Although almost half of the surveyed total protected area contained genetic data from at least one local population, the proportion of protected area dropped precipitously when increasing the number to ≥ 5 local populations (24.2%), or ≥ 10 (15.2%; Table S2). The available genetic data is also unevenly distributed across the globe. While Europe and North America are data-rich (both continents harbor 72.17% of all genotyped terrestrial local populations), only 13% of the total area protected in Africa and 16% of area protected in Asia had one or more local populations surveyed (Fig. 1, Table S2). More than 90% of the area that is protected within the Southeast Pacific, Chile, New Zealand and Arctic Sea Marine realms (classified following 11; Fig. 1) did not have one population sampled (Fig. 1, Table S2).

We further identified major taxonomic gaps, as genetic information was (on average) available for less than 0.5% of the total number of species for many major taxonomic groups, with large variation among groups (Fig. 2). For instance, Insecta (a major declining taxonomic group containing nearly a million known species, *12*), has less than 0.015% of genetically-informed species, whereas ~6% of mammal species are genetically-informed. Notably, even the most data-rich taxonomic groups were represented by <10% of all their described species

(Fig. 2). Filling taxonomic and spatial gaps is important, as patterns of genetic variation may not be similar nor associated with the same drivers across taxa and regions (*6, 9, 13*).

Despite major gaps, locally, some PAs harbored a large amount of genetic data, which may be sufficient to start assessing their protection of genetic variation. For instance, the Channel Islands National Park, CA, USA (135 local populations, 30 species), the World Heritage Sites of the Galapagos Islands, Ecuador (278 local populations, 47 species) and the Great Barrier Reef, Australia (83 local populations, 23 species) are particularly data-rich (Fig. 1; Supplementary data 1).We also noted that a relatively high proportion of samples were collected in PAs (39%) (Figs. 1-2) relative to the overall global coverage of PAs (17%). This was especially true for some taxonomic groups: Polypodiophyta, Ctenophora and Phaeophyta each had more than 70% of their genetic data collected within PAs. PAs may be a focus for genetic data collection for a number of reasons, including that they contain target populations to sample, provide better access to natural or conserved areas of interest for conducting genetic-based studies, or because they may facilitate comprehensive or long-term research projects. Furthermore, ~42.75% of local populations with genetic data were located in hotspots for species biodiversity (see "Materials and Methods" section in the Supplementary Materials) and could be assessed to investigate PA design in these biodiversity-rich areas (e.g., in the North American Coastal Plain, the Mediterranean Basin and the California Floristic Province; Table S3).

Our results show that the global distribution of available nuclear genetic data at the population level is sparse and unevenly distributed within PAs and in other parts of the world, which could be addressed by the use of multiple approaches to gather and integrate genetic variation in biodiversity protection and monitoring efforts. First, in data-rich regions, data can already begin to be repurposed to assess genetic variation and connectivity levels within and outside of PAs (e.g., *6*), and to designate new PAs that integrate genetic criteria. Additionally, these data can set contemporary genetic baselines for many populations and species, facilitating temporal genetic monitoring programs (*14*). Second, to bridge global data gaps, genetic data could be correlated to geographic, environmental, and taxonomic (e.g., species diversity) surrogates to identify informative proxies at different spatial scales (*13*). Cohesive analyses of genetic data from many species at once (i.e., macrogenetic analyses, reviewed in *9*) can be powerful and effective in identifying drivers of genetic diversity and connectivity, and biotic/abiotic correlates such as the impact of PA design (e.g., size, shape, connectivity to other PAs). Third, in the absence of genetic information, we recommend the development and use of genetic indicators that can be estimated without genetic data (e.g., KMGBF headline indicator A.4, *1*). Fourth, additional capacity to collect genetic data, and to increase efforts to gather, harmonize and share existing genetic data can help to address large regional and taxonomic data gaps (*9*), particularly when focused in data-poor regions. Regarding this last objective, we found very low levels of data duplication (i.e., data included in more than one database) across the databases we examined (14%), even though at least some of these were targeting similar taxa and regions. Limited data overlap suggests that there is large potential to add more data from existing resources, including publications and open data repositories, if these data can be found and enriched with the necessary metadata to facilitate their reuse (e.g., geographical coordinates; *15*). Adopting universal reporting and archiving standards for

genetic and genomic data could improve the usability of any new data for monitoring progress toward global conservation targets (*15*).

Conclusions

Area-based conservation initiatives, like 30x30, offer the chance to increase biodiversity protection (including genetic diversity protection) in the next seven years and beyond. However, the availability of genetic data to assess and inform PA design and efficacy is still sparse across the globe and for many taxonomic groups. Targeted efforts can help to fill these gaps. In some regions, there is potential to inform efforts to improve the protection of intraspecific genetic variation and connectivity within and between PAs. Integrating genetics in PA assessments could help protect all biodiversity components and streamline and optimize efforts to meet multiple goals and targets of regional, national and global commitments on biodiversity conservation (KMGBF Targets 3 and 4; *7*). Overall, increased integration of intraspecific genetic data with spatial conservation planning tools (*4*, *5*) can help to ensure comprehensive conservation of myriad species and populations, and their unique genetic variation, in efforts to increase PAs.

References and Notes

1. S. Hoban, J. M. Da Silva, A. Mastretta-Yanes, C. E. Grueber, M. Heuertz, M. E. Hunter, J. Mergeay, I. Paz-Vinas, K. Fukaya, F. Ishihama, R. Jordan, V. Köppä, M. C. Latorre-Cárdenas, A. J. MacDonald, V. Rincon-Parra, P. Sjögren-Gulve, N. Tani, H. Thurfjell, L. Laikre, Monitoring status and trends in genetic diversity for the Convention on Biological Diversity: An ongoing assessment of genetic indicators in nine countries. *Conservation Letters.* e12953 (2023).

2. D. M. Leigh, A. P. Hendry, E. Vázquez-Domínguez, V. L. Friesen, Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications*. **12**, 1505–1512 (2019).

3. M. Exposito-Alonso, T. R. Booker, L. Czech, L. Gillespie, S. Hateley, C. C. Kyriazis, P. L. M. Lang, L. Leventhal, D. Nogues-Bravo, V. Pagowski, M. Ruffley, J. P. Spence, S. E. Toro Arana, C. L. Weiß, E. Zess, Genetic diversity loss in the Anthropocene. *Science*. **377**, 1431–1435 (2022).

4. M. Andrello, C. D'Aloia, A. Dalongeville, M. A. Escalante, J. Guerrero, C. Perrier, J. P. Torres-Florez, A. Xuereb, S. Manel, Evolving spatial conservation prioritization with intraspecific genetic data. *Trends in Ecology & Evolution*. **37**, 553–564 (2022).

5. E. S. Nielsen, J. O. Hanson, S. B. Carvalho, M. Beger, R. Henriques, F. Kershaw, S. Von Der Heyden, Molecular ecology meets systematic conservation planning. *Trends in Ecology & Evolution*. **38**, 143–155 (2023).

6. L. Figuerola-Ferrando, A. Barreiro, I. Montero-Serra, M. Pagès-Escolà, J. Garrabou, C. Linares, J. Ledoux, Global patterns and drivers of genetic diversity among marine habitat-forming species. *Global Ecol Biogeogr*. **32**, 1218–1229 (2023).

7. CBD, Decision adopted by the COP to the CBD 15/4. Kunming-Montreal Global Biodiversity Framework (2022), (available at https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-04-en.pdf).

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

9. D. M. Leigh, C. B. van Rees, K. L. Millette, M. F. Breed, C. Schmidt, L. D. Bertola, B. K. Hand, M. E. Hunter, E. L. Jensen, F. Kershaw, L. Liggins, G. Luikart, S. Manel, J. Mergeay, J. M. Miller, G. Segelbacher, S. Hoban, I. Paz-Vinas, Opportunities and challenges of macrogenetic studies. *Nat Rev Genet*. **22**, 791–807 (2021).

10. S. L. Maxwell, V. Cazalis, N. Dudley, M. Hoffmann, A. S. L. Rodrigues, S. Stolton, P. Visconti, S. Woodley, N. Kingston, E. Lewis, M. Maron, B. B. N. Strassburg, A. Wenger, H. D. Jonas, O. Venter, J. E. M. Watson, Area-based conservation in the twenty-first century. *Nature*. **586**, 217–227 (2020).

11. M. J. Costello, P. Tsai, P. S. Wong, A. K. L. Cheung, Z. Basher, C. Chaudhary, Marine biogeographic realms and species endemicity. *Nat Commun*. **8**, 1057 (2017).

12. R. Van Klink, D. E. Bowler, K. B. Gongalsky, A. B. Swengel, A. Gentile, J. M. Chase, Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*. **368**, 417–420 (2020).

13. C. Schmidt, S. Hoban, W. Jetz, Conservation macrogenetics: harnessing genetic data to meet conservation commitments. *Trends in Genetics*, S0168952523001816 (2023).

14. R. D. Clark, K. A. Catalano, K. S. Fitz, E. Garcia, K. E. Jaynes, B. N. Reid, A. Sawkins, A. A. Snead, J. C. Whalen, M. L. Pinsky, The practice and promise of temporal genomics for measuring evolutionary responses to global change. *Molecular Ecology Resources*, 1755–0998.13789 (2023).

15. D. M. Leigh, A. Vandergast, M. Hunter, E. Crandall, W. Funk, C. Garroway, S. Hoban, S. Oyler-McCance, C. Rellstab, G. Segelbacher, C. Schmidt, E. Vázquez-Domínguez, I. Paz-Vinas, Don’t make genetic data disposable: Best practices for genetic and genomic data archiving (preprint, EcoEvorXiv, 2023), doi:10.32942/X29025.

Supplementary Material references

16. S. Blanchet, J. G. Prunier, H. De Kort, Time to go bigger: Emerging patterns in macrogenetics. *Trends in Genetics*. **33**, 579–580 (2017).

17. C. M. French, L. D. Bertola, A. C. Carnaval, E. P. Economo, J. M. Kass, D. J. Lohman, K. A. Marske, R. Meier, I. Overcast, A. J. Rominger, P. P. A. Staniczenko, M. J.

Hickerson, Global determinants of insect mitochondrial genetic diversity. *Nat Commun*. **14**, 5276 (2023).

18. L. Dapporto, M. Menchetti, R. Vodă, C. Corbella, S. Cuvelier, I. Djemadi, M. Gascoigne-Pees, J. C. Hinojosa, N. T. Lam, M. Serracanta, G. Talavera, V. Dincă, R. Vila, The atlas of mitochondrial genetic diversity for Western Palaearctic butterflies. *Global Ecol Biogeogr*. **31**, 2184–2190 (2022).

19. K. L. Millette, V. Fugère, C. Debyser, A. Greiner, F. J. J. Chain, A. Gonzalez, No consistent effects of humans on animal genetic diversity worldwide. *Ecology Letters*. **23**, 55–67 (2020).

20. S. Manel, P.-E. Guerin, D. Mouillot, S. Blanchet, L. Velez, C. Albouy, L. Pellissier, Global determinants of freshwater and marine fish genetic diversity. *Nature Communications*. **11**, 692 (2020).

21. S. Theodoridis, D. A. Fordham, S. C. Brown, S. Li, C. Rahbek, D. Nogues-Bravo, Evolutionary history and past climate change shape the distribution of genetic diversity in terrestrial mammals. *Nat Commun*. **11**, 2557 (2020).

22. M. C. Baranzelli, A. Cosacov, S. M. Sede, M. V. Nicola, A. N. Sérsic, Anthropocene refugia in Patagonia: A macrogenetic approach to safeguarding the biodiversity of flowering plants. *Biological Conservation*. **268**, 109492 (2022).

23. P. Sunnucks, Efficient genetic markers for population biology. *Trends in Ecology & Evolution*. **15**, 199–203 (2000).

24. F. W. Allendorf, Genetics and the conservation of natural populations: allozymes to genomes. *Mol Ecol*. **26**, 420–430 (2017).

25. M. Kardos, E. E. Armstrong, S. W. Fitzpatrick, S. Hauser, P. W. Hedrick, J. M. Miller, D. A. Tallmon, W. C. Funk, The crucial role of genome-wide genetic variation in conservation. *Proc Natl Acad Sci USA*. **118**, e2104642118 (2021).

26. I. Paz-Vinas, E. L. Jensen, L. D. Bertola, M. F. Breed, B. K. Hand, M. E. Hunter, F. Kershaw, D. M. Leigh, G. Luikart, J. Mergeay, J. M. Miller, C. B. Van Rees, G. Segelbacher, S. Hoban, Macrogenetic studies must not ignore limitations of genetic markers and scale. *Ecology Letters*. **24**, 1282–1284 (2021).

27. C. Schmidt, C. J. Garroway, The conservation utility of mitochondrial genetic diversity in macrogenetic research. *Conserv Genet* (2021), doi:10.1007/s10592-021-01333-6.

28. E. R. Lawrence, J. N. Benavente, J.-M. Matte, K. Marin, Z. R. R. Wells, T. A. Bernos, N. Krasteva, A. Habrich, G. A. Nessel, R. A. Koumrouyan, D. J. Fraser, Geo-referenced population-specific microsatellite data across American continents, the MacroPopGen Database. *Scientific Data*. **6**, 14 (2019).

29. E. R. Lawrence, J. N. Benavente, Jean-Michel Matte, K. Marin, Z. Wells, T. A. Bernos, N. Krasteva, A. Habrich, G. Nessel, Ramela Arax Koumrouyan, D. J. Fraser, MacroPopGen Database: Geo-referenced population-specific microsatellite data across the American continents (2019), doi:10.6084/M9.FIGSHARE.7207514.V1.

30. J. Beninde, E. M. Toffelmier, A. Andreas, C. Nishioka, M. Slay, A. Soto, J. P. Bueno, G. Gonzalez, H. V. Pham, M. Posta, J. L. Pace, H. B. Shaffer, CaliPopGen: A genetic and life history database for the fauna and flora of California. *Sci Data*. **9**, 380 (2022).

31. H. B. Shaffer, J. Beninde, E. M. Toffelmier, A. Andreas, C. Nishioka, M. Slay, A. Soto, J. P. Bueno, H. V. Pham, G. Gonzalez, J. L. Pace, CaliPopGen: A genetic and life history database for the fauna and flora of California (2022), doi:10.6084/M9.FIGSHARE.C.5518389.V1.

32. C. Schmidt, M. Domaratzki, R. P. Kinnunen, J. Bowman, C. J. Garroway, Continent-wide effects of urbanization on bird and mammal genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*. **287**, 20192497 (2020).

33. C. Schmidt, C. J. Garroway, The population genetics of urban and rural amphibians in North America. *Molecular Ecology*. **30**, 3918–3929 (2021).

34. C. Schmidt, C. J. Garroway, Systemic racism alters wildlife genetic diversity. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2102860119 (2022).

35. E. Karachaliou, C. Schmidt, C. J. Garroway, A global macrogenetic database for marine fish. Metadata included in Supplementary data 2 (this paper).

36. R. Clark, M. Pinsky, Global patterns of genetic diversity in marine fishes (preprint, Authorea, 2023), doi:10.22541/au.168726266.68173061/v1 (2023).

37. H. De Kort, J. G. Prunier, S. Ducatez, O. Honnay, M. Baguette, V. M. Stevens, S. Blanchet, Life history, climate and biogeography interactively affect worldwide genetic diversity of plant and animal populations. *Nature Communications*. **12**, 516 (2021).

38. H. D. Kort, J. G. Prunier, S. Ducatez, O. Honnay, M. Baguette, V. M. Stevens, Blanchet Simon, Data and scripts Genetic Diversity (2020), doi:10.6084/M9.FIGSHARE.13373363.

39. E. D. Crandall, R. H. Toczydlowski, L. Liggins, A. E. Holmes, M. Ghoojaei, M. R. Gaither, B. E. Wham, A. L. Pritt, C. Noble, T. J. Anderson, R. L. Barton, J. T. Berg, S. G. Beskid, A. Delgado, E. Farrell, N. Himmelsbach, S. R. Queeno, T. Trinh, C. Weyand, A. Bentley, J. Deck, C. Riginos, G. S. Bradburd, R. J. Toonen, Importance of timely metadata curation to the global surveillance of genetic diversity. *Conservation Biology*. **37**, e14061 (2023).

40. UNEP-WCMC, IUCN, Protected Planet: The World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures

(WD-OECM) [Online]. *Cambridge, UK: UNEP-WCMC and IUCN* (2022) (available at https://www.protectedplanet.net/).

41. J. O. Hanson, wdpar: Interface to the World Database on Protected Areas. *JOSS*. **7**, 4594 (2022).

42. R Development Core Team, R: A language and environment for statistical computing. *Vienna, Austria, R Foundation for Statistical Computing.* (2022).

43. S. H. M. Butchart, M. Clarke, R. J. Smith, R. E. Sykes, J. P. W. Scharlemann, M. Harfoot, G. M. Buchanan, A. Angulo, A. Balmford, B. Bertzky, T. M. Brooks, K. E. Carpenter, M. T. Comeros-Raynal, J. Cornell, G. F. Ficetola, L. D. C. Fishpool, R. A. Fuller, J. Geldmann, H. Harwell, C. Hilton-Taylor, M. Hoffmann, A. Joolia, L. Joppa, N. Kingston, I. May, A. Milam, B. Polidoro, G. Ralph, N. Richman, C. Rondinini, D. B. Segan, B. Skolnik, M. D. Spalding, S. N. Stuart, A. Symes, J. Taylor, P. Visconti, J. E. M. Watson, L. Wood, N. D. Burgess, Shortfalls and Solutions for Meeting National and Global Conservation Area Targets. *Conservation letters*. **8**, 329–337 (2015).

44. C. A. Runge, J. E. M. Watson, S. H. M. Butchart, J. O. Hanson, H. P. Possingham, R. A. Fuller, Protected areas and global conservation of migratory birds. *Science*. **350**, 1255–1258 (2015).

45. P. Visconti, M. Di Marco, J. G. Álvarez-Romero, S. R. Januchowski-Hartley, R. L. Pressey, R. Weeks, C. Rondinini, Effects of Errors and Gaps in Spatial Data Sets on Assessment of Conservation Progress: Errors and Gaps in Spatial Data Sets. *Conservation Biology*. **27**, 1000–1010 (2013).

46. N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities. *Nature*. **403**, 853–858 (2000).

47. R. A. Mittermeier, W. R. Turner, F. W. Larsen, T. M. Brooks, C. Gascon, "Global Biodiversity Conservation: The Critical Role of Hotspots" in *Biodiversity Hotspots*, F. E. Zachos, J. C. Habel, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011; https://link.springer.com/10.1007/978-3-642-20992-5_1), pp. 3–22.

48. R. F. Noss, W. J. Platt, B. A. Sorrie, A. S. Weakley, D. B. Means, J. Costanza, R. K. Peet, How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. *Diversity Distrib.* **21**, 236–244 (2015).

Acknowledgements: We thank Dr. Sean Hoban, Dr. Rachel Toczydlowski and Dr. Robert Fisher for their insightful comments and suggestions.

Funding: I.P.-V. was supported by the U.S. Geological Survey (USGS) Powell Center for Synthesis and Analysis. This work was conducted as a part of the "Standardizing, Aggregating, Analyzing and Disseminating Global Wildlife Genetic and Genomic Data for Improved Management and Advancement of Community Best Practices Working Group"

supported by the John Wesley Powell Center for Analysis and Synthesis, funded by the USGS. RC and MP were supported by US National Science Foundation #OISE-1743711.

Author contributions:
Conceptualization: All
Data curation: IP-V, EDC, RDC, CS, EK, HdK
Formal Analysis: IP-V, AGV
Funding acquisition: AGV, MEH
Methodology: IP-V, AGV
Project administration: IP-V, AGV, MEH
Resources: EDC, RDC, MLP, CJG, CS, EK, SB, HdK, JF
Visualization: IP-V, AGV, CS
Writing – original draft: IP-V, AGV, CS, DML, MEH
Writing – review & editing: All

Competing interests: Authors declare that they have no competing interests.

Disclaimer: Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Data and materials availability: Supplementary data files 1 and 2 will be deposited in an open repository once the paper will be accepted for publication in a journal.

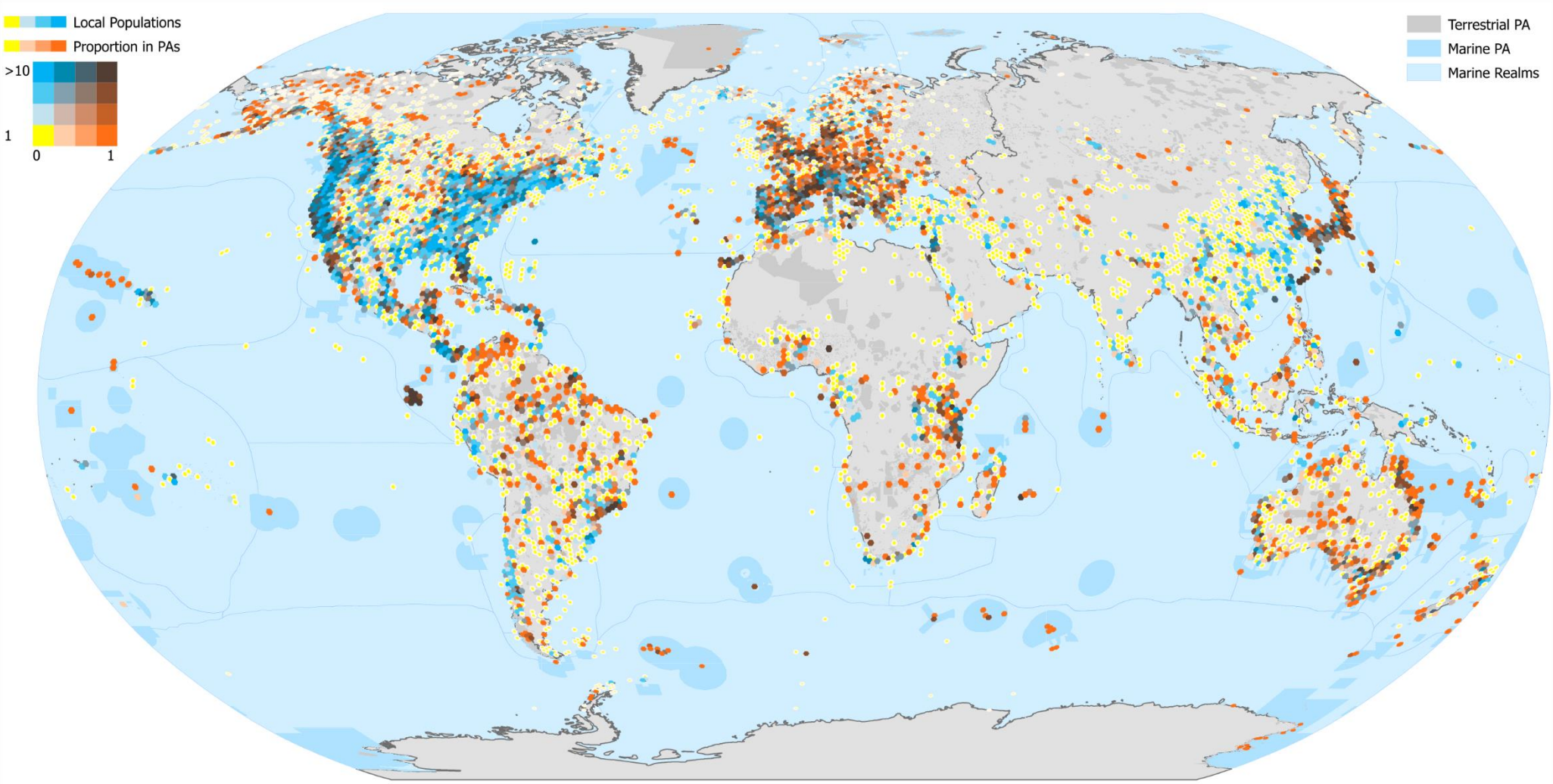

Fig. 1: **Global map of population-level nuclear genetic data available within and outside marine and terrestrial protected areas (PAs)**. Vertical color-scale (yellow to blue) represents the number of local populations within 8,000 km^2 hexagons, while the horizontal color-scale (yellow to orange) represents the proportion of local populations within the hexagon inside a PA. Yellow hexagons have just one local population that is outside PAs, while dark brown hexagons have >10 local populations all within PAs. Marine realms have been delimited following (*11*). Terrestrial and marine PAs have been defined according to the World Database on Protected Areas (accessed 23 November 2022).

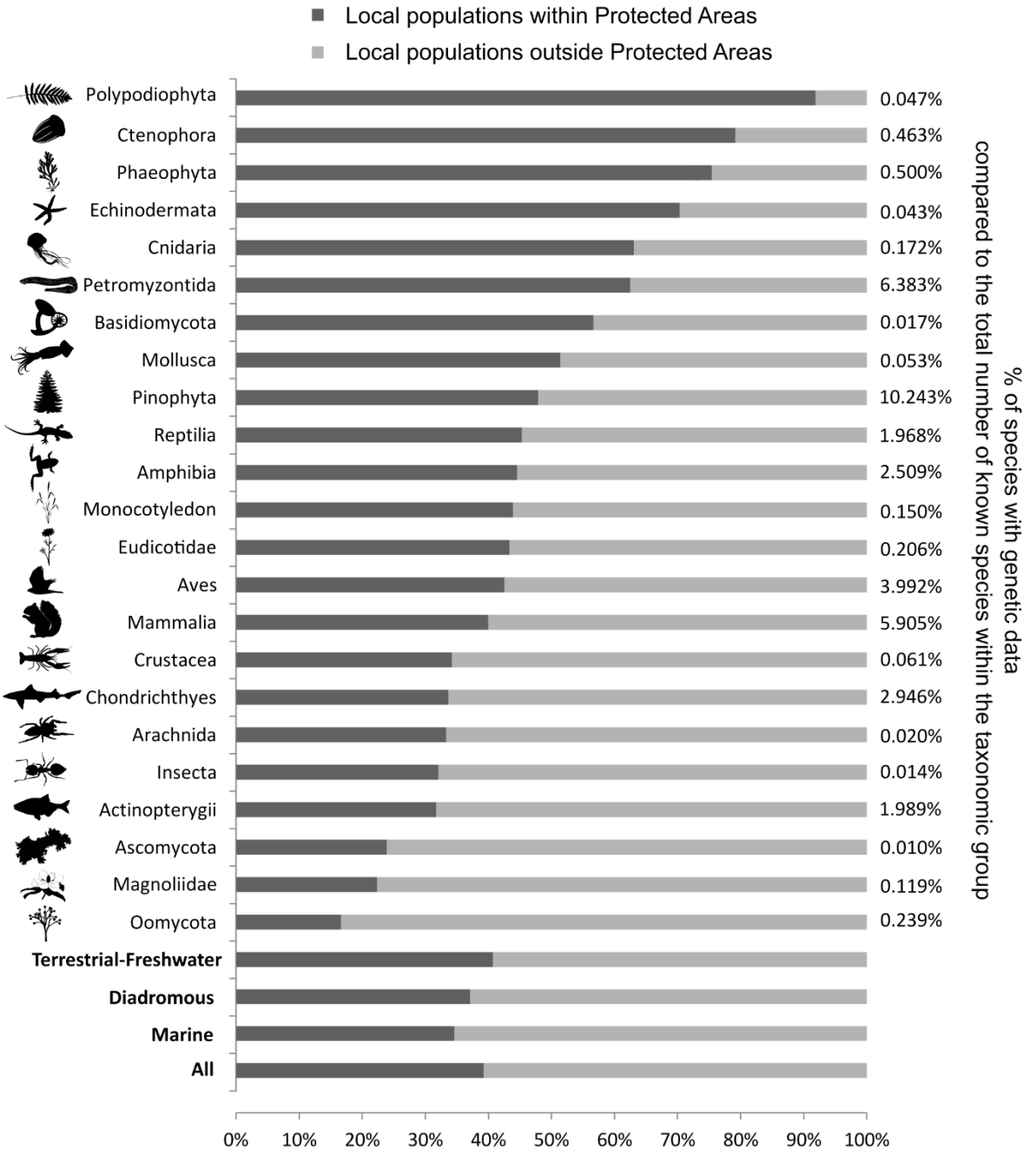

Fig. 2: **Proportion of local populations within and outside protected areas**. Proportions are displayed *per* taxonomic group and *per* species' primary habitat type (at bottom in bold). Results for taxonomic groups with fewer than 20 local populations in the meta-database have not been represented, but have been included in the categories in bold. Percentages at the right of the figure represent the proportion of species with genetic data compared to the total number of known species within the taxonomic group.

Supplementary Materials

Materials and Methods

Building the meta-database of nuclear genetic data

To gather geographic coordinates from organisms that have been genotyped at nuclear loci worldwide, we built a meta-database of nuclear genetic data previously compiled in seven macrogenetic databases (combined data in Supplementary data 2). Macrogenetic databases gather summaries of genetic variation reported in publications (i.e., Class II databases; *9*) or raw genetic data (i.e., genotypes) from public repositories like Dryad or Figshare (i.e., Class III databases; *9*), and/or metadata associated with biological samples stored in laboratories (e.g., geographic coordinates of samples and populations). By repurposing genetic variation data at large taxonomic, temporal and/or spatial scales, macrogenetic databases have recently unlocked the study of large-scale patterns and drivers of intraspecific genetic variation (*16*, *9*). Many macrogenetic studies have focused on mitochondrial (e.g., *17–21*) or chloroplastic (*22*) DNA data. We, however, only included metadata from macrogenetic databases that focused on population-level nuclear genetic data (i.e., microsatellites, AFLPs, SNPs), given the long-standing consensus that nuclear genotypic data are the most appropriate for tracking fine-scale population processes over short timeframes (*23–25*), and the many issues that have been highlighted concerning the use of organelle-based DNA, such as mtDNA loci data for estimating whole-genome genetic variation and for macrogenetic analyses (*9*, *26*, *27*).

We extracted geographic coordinates from the macrogenetic databases (Table S1). The first database (i.e., "Macropopgen" database; *28*, *29*) is a class II database containing multiple genetic variation estimates measured at the site level using microsatellite loci for 897 vertebrate species (terrestrial vertebrates and freshwater fish) across North and South America (Table S1). The second, Calipopgen (*30*, *31*; Class II) contains genetic data and life-history data for populations from multiple eukaryotic species located within marine and terrestrial ecoregions of California (only data using nuclear markers was used here). The third is a population-level database compiled by Schmidt et al. (*32*, *33*, *34*; Class III) containing microsatellite genotypes of North American terrestrial vertebrate species extracted from public repositories (Table S1). The fourth is an extension of the Schmidt et al. database that compiled genotypic data for marine fishes worldwide (*35*; Class III). The fifth database, compiled by Clark & Pinsky (*36*; Class II), also focused on marine fish worldwide, but was based on site-level georeferenced genetic estimates extracted from the literature. The sixth was a database compiled by De Kort et al. (*37*, *38*; Class II) containing population-level AFLP and microsatellite data for plant, terrestrial vertebrate and mollusk species worldwide (Table S1). The seventh was the metadata database built by Crandall et al. (*39*), which contains metadata including geographical coordinates for accessions to the International Nucleotide Sequence Database Collaboration's Sequence Read Archive (INSDC SRA) retrieved during a datathon for >400 genomic datasets comprising >40,000 individuals from >700 species from multiple taxonomic groups (Table S1).

For each dataset, we extracted information on the species, its taxonomic level or group, geographic coordinates of sampling locations for groups of genotyped individuals

(which we designate as "local populations" following *33*), numbers of sampled individuals per local population (when available), and other study-related metadata (i.e., first author, author list or DOI identifier of the original study, when available). Study-related metadata fields were mostly used to identify, through cross-filtering procedures, datasets, studies and/or populations being duplicated among different macrogenetic databases. These databases differ in terms of spatial scope (regional vs. worldwide) and in the procedures used by their authors to georeference and delimit populations. Thus, when duplicated studies or datasets were identified among databases, we always retained information from the database that extracted the highest number of georeferenced local populations from the study, as it was infeasible to check manually which local populations were best representing the true delimitations of biological populations of the target species. In total, we found a total of 13.97% of duplicated data, and the original combined dataset with 38,284 local populations was reduced to 36,356 unique (non-duplicated) local populations for final analyses. Finally, we homogenized taxonomic group names of species among databases and identified whether their habitat of preference was marine, terrestrial (including freshwater habitats) or diadromous.

Protected areas dataset

The World database on protected areas (WDPA; *40*) is the most comprehensive global database of PAs, MPAs and OECMs. We used the 'wdpar' package (*41*) of the R Statistical software v 4.2.2 (*42*) to download and clean the global dataset of PAs (protectedplanet.net; downloaded 23 November 2022) following methods outlined in (*43*) and (*44*). Briefly, this included removing areas with "proposed" or unknown status, excluding UNESCO Biosphere reserves, buffering PAs represented as point localities to circular areas using their reported spatial extent (*45*), and repairing invalid geometries. Because many protected areas in the WDPA have overlapping areas, we transformed and flattened overlapping areas into a raster with 1 km² grid cells to make global area calculations.

Nuclear genetic data availability analyses

We assessed spatial and taxonomic gaps in data availability among species and taxonomic groups by comparing data from our meta-database with our refined protected areas database. We specifically used ArcGIS 10.7 tools (ESRI) to assess how many local populations were sampled inside and outside protected areas, depending on the taxonomic group of the species, and the type of habitat where species live. When using spatial selection and join procedures to select PAs containing populations with genetic data, we considered a 1 km buffer around local population coordinates to account for coordinate uncertainty due to the use of different georeferencing procedures by the authors of the original macrogenetic databases. One kilometer corresponds to the median coordinate uncertainty estimated for all INSDC Biosamples from the Crandall et al. metadata database (*39*). To estimate geographic coverage, we summed the area of PAs that contained at least one genetically-sampled local population and divided it by the total protected area within continents and marine realms (Table S2). To map global genetic data density, we aggregated the global distribution of genetic data using an optimized hotspots analysis of point counts within a hexagonal grid using default parameters in ArcGIS Pro 3.1.1, and calculated the proportion of points inside protected areas within each hexagonal cell (cell size approx. 8,000 km^2). Final layers were projected into World Robinson projection for visualization. To estimate taxonomic coverage,

we extracted data on the total number of described species per taxonomic group from the "Catalogue Of Life" COL web portal (https://www.catalogueoflife.org/, last accessed 2nd of July 2023) to estimate the proportion of species in the meta-database in relation to the total number of known species per taxonomic group (i.e., species coverage of the analysis).

Hotspots of biodiversity

We further assessed genetic data availability within and outside hotspots of biodiversity, by comparing data from our meta-database with spatial data from the Earth's 36 biodiversity hotspots (46–48) using ArcGIS 10.7 (spatial data downloaded 26 June 2023 from https://hub.arcgis.com/datasets/ba55aa1bff5447e7b72559b8dc1a0e83_0/about).

Table S1. **Details of the meta-database and the different macrogenetic databases used to build it.** The table provides a description and summary of the different macrogenetic databases used to extract and compile geographic coordinates data from multiple local populations across geographical and taxonomic scales in a single meta-database. The type of macrogenetic database (Class II: compilation of summaries of genetic variation reported in publications; Class III: compilation of raw genetic data from public repositories; *9*), the numbers of included species, local populations and genotyped individuals, and the types of habitat and geographical and taxonomic scopes of the macrogenetic databases are reported.

Macrogenetic database	*Type of macrogenetic database*	*Number of species*	*Number of local populations*	*Number of genotyped individuals*	*Geographical scale*	*Type of habitat*	*Taxonomic scope*
MACROPOPGEN (Lawrence et al. 2019; *28, 29*)	Class II	897	9,090	561,605	Americas	Terrestrial / Freshwater	Mammals, Birds, Amphibians, Reptiles, Bony fishes
Schmidt et al. 2020, 2021, 2022 (*32-34*)	Class III	219	2,824	96,195	USA, Canada	Terrestrial / Freshwater	Mammals, Birds, Amphibians, Reptiles
Karachaliou et al. (*35*)	Class III	73	1,136	74,615	Worldwide	Marine	Bony and cartilaginous fishes
Clark and Pinsky 2023 (*36*)	Class II	329	3,080	621,486	Worldwide	Marine	Bony and cartilaginous fishes
De Kort et al. 2021 (*37, 38*)	Class II	714	8,356	246,422	Worldwide	Terrestrial / Freshwater	Plants, Mammals, Birds, Amphibians, Reptiles, Molluscs
Crandall et al. 2023 (*39*)	Class III	730	10,289	42,444	Worldwide	Terrestrial / Freshwater / Marine	31 major taxonomic groups from four Kingdoms (Animalia, Plantae, Chromista, Fungi)
CALIPOPGEN (Beninde et al. 2022; *30, 31*)	Class II	286	3,711	690,544	California	Terrestrial / Freshwater / Marine	24 major taxonomic groups from four Kingdoms (Animalia, Plantae, Chromista, Fungi)
Total Meta-database (this study)	Geographical coordinates	2,809	36,356	2,016,527	Worldwide	Terrestrial / Freshwater / Marine	36 major taxonomic groups from four Kingdoms (Animalia, Plantae, Chromista, Fungi)

Table S2. Genetic data availability in protected areas per continent and Marine realm. Total protected area was estimated from the World Database on Protected Areas (accessed 23 November 2022).

Continents	**Total Protected Area (km2)**	Percent ≥ 1 sampled local population	Percent ≥ 5 sampled local population	Percent ≥ 10 sampled local population	**Total number of local populations sampled**
Asia	3,814,803	16.48%	2.47%	1.07%	3,339
North America	3,934,670	69.14%	40.85%	11.84%	14,736
Europe	1,710,929	28.52%	4.48	1.01%	4,494
Africa	5,493,782	12.83%	2.64%	0.36%	1,243
South America	4,587,639	35.50%	10.00%	3.30%	2,167
Oceania	105,168	32.17%	1.59%	0.62%	282
Australia	1,561,103	44.85%	3.76%	2.19%	916
Antarctica	44,732	96.07%	95.18%	94.42%	26
Total Terrestrial Area	**21,252,826**	**32.69%**	**11.69%**	**3.63%**	**27,203**
Marine Realms *(12)*					
1 Inner Baltic Sea	48,179	24.75%	7.12%	0.00%	64
2 Black Sea	23,483	11.77%	0.10%	0.00%	14
3 NE Atlantic	313,293	53.91%	26.26%	3.31%	428
4 Norwegian Sea	63,563	26.04%	0.00%	0.00%	77
5 Mediterranean	188,082	65.29%	40.31%	0.01%	422
6 Arctic seas	299,811	7.85%	4.69%	4.54%	38
7 N. Pacific	217,221	55.06%	33.31%	32.51%	3,127
8 N. American Boral	513,972	43.50%	18.76%	0.00%	29
9 Mid-tropical North Pacific Ocean	2,389,033	99.78%	63.29%	62.93%	46
10 South-east Pacific	1,967,188	0.00%	0.00%	0.00%	7
11 Caribbean & Gulf of Mexico	451,768	49.43%	19.42%	4.26%	817
12 Gulf of California	1,024,051	69.18%	17.24%	16.50%	186
13 Indo-Pacific seas & Indian Ocean	1,596,508	52.61%	0.66%	0.01%	301
14 Gulfs of Aqaba, Aden, Suez, Red Sea	26,680	28.66%	0.00%	0.00%	38

15 Tasman Sea	527,359	67.43%	50.80%	0.00%	5
16 Coral Sea	3,180,124	81.69%	65.27%	11.40%	199
17 Mid South Tropical Pacific	3,239,694	85.70%	0.00%	0.00%	72
18 Offshore & NW North Atlantic	1,614,851	47.83%	26.42%	11.37%	534
19 Offshore Indian Ocean	758,882	87.26%	67.42%	67.42%	7
20 Offshore W Pacific	1,184,827	32.91%	14.33%	11.92%	2,059
21 Offshore S Atlantic	2,199,411	7.11%	0.66%	0.14%	55
22 Offshore mid-E pacific	110,526	89.28%	0.00%	0.00%	6
23 Gulf of Guinea	71,461	51.50%	0.00%	0.00%	12
24 Rio de La Plata	110,989	25.74%	0.00%	0.00%	39
25 Chile	86,396	7.87%	0.00%	0.00%	29
26 S Australia	521,427	17.35%	0.04%	0.00%	42
27 S. Africa	71,454	25.44%	0.00%	0.00%	99
28 New Zealand	1,039,588	3.71%	0.00%	0.00%	55
29 NW Pacific	31,783	23.90%	21.29%	21.29%	254
30 Southern Ocean	5,332,852	81.39%	77.01%	73.57%	92
Total Marine Area	**29,204,456**	**59.00%**	33.25%	**23.69%**	**9,153**
Total Global Area	**50,457,282**	**47.92%**	24.17%	**15.24%**	**36,356**

Table S3. **Number of local populations inside protected areas within biodiversity hotspots area.** Biodiversity hotspots are sorted by ascending number of local populations inside PAs within the hotspot area

Hotspot Area	**Number of local populations inside PAs within the hotspot area**
East Melanesian Islands	0
Horn of Africa	3
Western Ghats and Sri Lanka	3
Mountains of Central Asia	7
Irano-Anatolian	8
Caucasus	11
Cape Floristic Region	11
New Caledonia	11
Succulent Karoo	13
Mountains of Southwest China	16
Himalaya	19
Guinean Forests of West Africa	20
Maputaland-Pondoland-Albany	20
Wallacea	21
Sundaland	44
Madagascar and the Indian Ocean Islands	45

Tropical Andes	55
Cerrado	56
Chilean Winter Rainfall and Valdivian Forests	60
Indo-Burma	67
Forests of East Australia	67
Coastal Forests of Eastern Africa	68
Madrean Pine-Oak Woodlands	95
New Zealand	98
Southwest Australia	102
Polynesia-Micronesia	105
Eastern Afromontane	106
Caribbean Islands	140
Atlantic Forest	150
Tumbes-Choco-Magdalena	306
Philippines	352
Mesoamerica	464
Japan	485
North American Coastal Plain	538
Mediterranean Basin	1,225
California Floristic Province	1,886