Article

Global impoverishment of natural vegetation revealed by dark diversity

https://doi.org/10.1038/s41586-025-08814-5

Received: 7 February 2024

Accepted: 19 February 2025

Published online: 2 April 2025

Open access

Check for updates

Anthropogenic biodiversity decline threatens the functioning of ecosystems and the many benefits they provide to humanity[1]. As well as causing species losses in directly affected locations, human influence might also reduce biodiversity in relatively unmodified vegetation if far-reaching anthropogenic effects trigger local extinctions and hinder recolonization. Here we show that local plant diversity is globally negatively related to the level of anthropogenic activity in the surrounding region. Impoverishment of natural vegetation was evident only when we considered community completeness: the proportion of all suitable species in the region that are present at a site. To estimate community completeness, we compared the number of recorded species with the dark diversity—ecologically suitable species that are absent from a site but present in the surrounding region[2]. In the sampled regions with a minimal human footprint index, an average of 35% of suitable plant species were present locally, compared with less than 20% in highly affected regions. Besides having the potential to uncover overlooked threats to biodiversity, dark diversity also provides guidance for nature conservation. Species in the dark diversity remain regionally present, and their local populations might be restored through measures that improve connectivity between natural vegetation fragments and reduce threats to population persistence.

Direct detrimental effects of anthropogenic activity on the biodiversity of natural ecosystems have been extensively documented[3,4]. For example, conversion of natural forest into urban landcover[5] or transformation of grassland into cropland[6] causes conspicuous declines in biodiversity. Biodiversity may also decline in ecosystems that are not directly modified but occur in regions in which human activities have caused habitat fragmentation[7] or exert diffuse effects on natural areas—through pollution, for example[8]. Although compelling case studies show the influence of human activities on surrounding natural vegetation, beyond a direct area of impact[8–10], there is no empirical evidence demonstrating the generality of regional-scale anthropogenic effects on local biodiversity in natural vegetation. Comparisons of relatively undisturbed vegetation inside and outside protected areas have revealed no discernible differences in local biodiversity[11], but this overlooks the possibility that biodiversity has declined systematically in both settings[12,13]. The lack of empirical evidence might stem from the masking effect of high variation in biodiversity across regions and along ecological gradients[14–16]. We hypothesize that anthropogenic impoverishment of natural ecosystems can be revealed by the dark diversity—species that are ecologically suitable and present in a region but currently absent from a given site[2]. Dark diversity allows estimation of community completeness, a biodiversity metric that represents the proportion of all suitable species in a region that are actually present at a site[17]. This metric is globally comparable because it accounts for natural variation in potential biodiversity. Estimating the ecological suitability of species that are absent from a site is challenging, but methodological advances offer a solution based on species co-occurrences[18].

The notion of dark diversity aligns with Whittaker's classic alpha–beta–gamma diversity framework[19]—a cornerstone of modern biodiversity research (Fig. 1). In Whittaker's work, alpha diversity represented the number of species at a particular site, gamma diversity comprised all species found in the surrounding region and beta diversity described changes in community composition along environmental gradients. The dark diversity concept is taxon-oriented, because it considers the suitability of each absent species for a study site. When aggregated, alpha and dark diversity together constitute the site-specific species pool, which includes only those species from the region that are suitable for a given site on the basis of its ecological conditions. In this context, beta diversity, as first defined by Whittaker, can be articulated as the change in site-specific species pools within gamma diversity. This is sometimes referred to as 'structured' beta diversity, whereas 'unstructured' beta diversity represents the variation in species composition among sampled sites within an ecologically similar area[20,21]. The dark diversity concept enhances the alpha–beta–gamma framework by providing a site-specific toolbox that complements alpha diversity at a site with the set of suitable yet absent species (dark diversity), the biodiversity potential of the site (species pool size) and the degree to which this potential is realized (community completeness).

Alpha diversity is the most commonly used biodiversity metric, but it depends on variation in natural biodiversity potential between regions (for example, boreal versus temperate regions; North America versus East Asia) and ecological conditions within regions (for example, wetlands versus forests; south-facing versus north-facing slopes). Speciation, large-scale dispersal, species sorting and stochastic variation have produced site-specific species pools of considerably different sizes[22]. Community completeness accounts for such variation by quantifying the extent to which the biodiversity potential (that is, the site-specific species pool) is realized locally[17]. Even in natural ecosystems, some

A list of authors and their affiliations appears at the end of the paper.

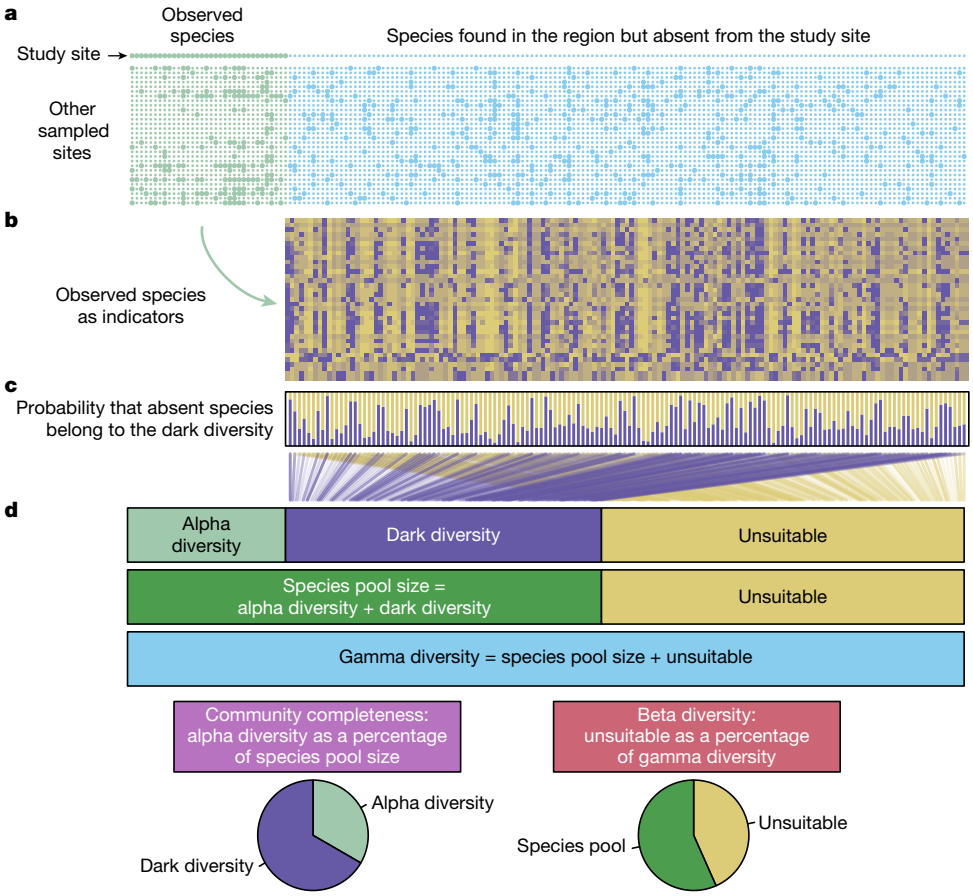


Fig. 1 | Estimating dark diversity and related biodiversity metrics in ecological communities. a, Data included a local study site where certain species were present, but many species sampled elsewhere in the region were absent. To estimate the probability that a species that is absent from the site but present in the region belongs to the dark diversity of the site, we used information about species co-occurrences at other sites in the region. **b**, We calculated an indicator matrix in which each present species indicated the ecological suitability of each absent species for the study site. We compared the observed number of co-occurrences with the number of co-occurrences expected at random (according to the hypergeometric distribution) and standardized the difference using the standard deviation from the hypergeometric distribution. **c**, By averaging across all observed species, each absent species was assigned a probability of belonging to the dark diversity for the study site. Consequently, the dark diversity was a fuzzy set to which species belonged to varying degrees. **d**, Several biodiversity metrics were characterized for each site in the region. Alpha diversity was the number of species recorded at the site, and gamma diversity was the total number of species recorded in a region. The size of dark diversity was estimated as the sum of the probabilities of absent species belonging to the dark diversity of the study site. Alpha and dark diversity together formed the site-specific species pool, and gamma diversity not falling into this category was considered the unsuitable part of gamma diversity; that is, belonging to the species pools of other sites. We investigated the percentage of the species pool that was present among the alpha diversity (community completeness) and the turnover of species pools in the region, expressed as the percentage of gamma diversity that was unsuitable for the study site (beta diversity).

suitable taxa might be absent owing to natural processes that cause local extinction or limit recolonization. Such limiting processes vary along environmental gradients, reflected in the global patterns of plant persistence strategies[23] and interactions with other organisms; for example, seed predators[24]. Consequently, there is likely to be natural variation in community completeness across broad environmental gradients[25]. In addition, regions with high geodiversity or a mosaic of vegetation types (that is, high structured beta diversity) might have lower community completeness because the isolation of natural habitat fragments and the likelihood of local extinction increase[26]. Furthermore, climatic conditions leave some regions prone to extreme events, such as natural fire, that cause local species loss[27,28]. Nevertheless, in addition to natural variation, human activities might strongly influence community completeness by reducing the persistence of local populations; for example, by promoting highly competitive taxa (through eutrophication, for instance[8]) or by restricting mutualistic interactions (reducing pollinators, for instance[29]). Similarly, human activities might hinder the recolonization of suitable sites through habitat fragmentation[7] and loss of seed-dispersing animals[30].

To determine whether anthropogenic impoverishment of natural vegetation is a worldwide phenomenon, we established DarkDivNet, a global collaborative research network[31]. Using a standardized methodology, we assessed both the alpha and the dark diversity of vascular plants across 5,415 sites with relatively intact natural or semi-natural vegetation, in 119 regions, spanning a wide range of vegetation types and representative of most global climatic conditions on all vegetated continents (Extended Data Figs. 1 and 2).

In our study, 'site' refers to a 100-m^2 area in which vegetation was sampled, and 'region' represents the surrounding area of approximately 300 km^2. Each region encompasses at least 30 sites, representing the natural and semi-natural vegetation typical of the region. We first confirmed that the sampling area of 100 m^2 provided highly similar estimates of dark diversity to those obtained from a considerably larger area of 2,500 m^2 (Extended Data Fig. 3). We assessed alpha diversity as the number of all vascular plant species found at each site. To estimate dark diversity, we used a fuzzy set approach in which all species occurring in the region but absent from the site were assigned a probability of inclusion in the dark diversity on the basis of an established

co-occurrence methodology[18]. The use of probabilities maximizes the amount of information used for estimating dark diversity. Specifically, co-occurrences were based on the species composition of 30 randomly selected sites in the region (Fig. 1a). Using a subset of regions in which 60 sites were available yielded highly similar outcomes, indicating that 30 sites were sufficient for estimating co-occurrence patterns among species (Extended Data Fig. 3). We estimated the degree to which each species present in a region but absent from a site co-occurred with species found at the site, and compared it with random expectation, mathematically described by the hypergeometric distribution (Fig. 1b). If an absent species co-occurred with a present species more than would be randomly expected, they probably shared ecological requirements, and the present species provided a positive indication of the site's suitability for the absent species. The overall suitability of the site for the absent species was estimated by averaging the suitability indications from all species present at the site (Fig. 1c). The magnitude of dark diversity at a site was then estimated as the sum of these suitability estimates (probabilities of absent species belonging to the dark diversity of the site, ranging between 0 and 1) across all absent species. The unsuitable fraction of gamma diversity reflects the species belonging to different site-specific species pools in the same region. Using alpha diversity, dark diversity and the unsuitable diversity found in the region, we calculated other biodiversity metrics for each site to have a full description of biodiversity (Fig. 1d): site-specific species pool size as the sum of alpha and dark diversity; gamma diversity as the total set of species found in a region (this value was the same for each site within a region); community completeness as the proportion of the site-specific species pool size represented by alpha diversity; and beta diversity as a quantification of the extent to which gamma diversity exceeds the site-specific species pool size (that is, the proportion of gamma diversity that is unsuitable for the study site and is more likely to be associated with different site-specific species pools in the region). In this way, we specifically quantified the 'structured' beta diversity, or turnover in site-specific species pool composition due to environmental gradients. In the statistical analyses, community completeness and beta diversity were included as log-ratios (logit transformation of percentages) to improve the distribution of the data. We used two independent datasets (expert assessments and examination of species found in the close vicinity of the site) to ensure that the co-occurrence method provided consistent estimates of species suitability for dark diversity (Extended Data Fig. 2). We also determined that, for this particular dataset, the hypergeometric method outperformed an alternative approach—joint species distribution modelling[32] (see Supplementary Methods).

The median community completeness of sites across all regions was 25% (95% confidence interval 15–46%), highlighting a frequent absence of suitable species despite their presence in surrounding regions (Fig. 2a). The existence of relatively high dark diversity is clearly a general phenomenon, but the large variation meant that sometimes much fewer species were present locally than might be expected from the specific site conditions. To understand how much variation in alpha diversity was explained by community completeness besides beta and gamma diversity, we used variation partitioning. We found that 33% (26–43%) of the variation in alpha diversity was explained by community completeness. Consequently, if human activities reduce the colonization and persistence of suitable species, resulting in lower community completeness, this could substantially affect alpha diversity. The largest proportion of variation in alpha diversity, 52% (40–61%), was explained by gamma diversity, reflecting the well-known match of local and regional diversity[33], whereas 14% (9–21%) was explained by beta diversity, reflecting how gamma diversity is distributed across different site-specific species pools. The strong dependence of alpha diversity on regional richness is clearly sufficient to mask the negative effect of human activities on alpha diversity.

We tested the hypothesis that impoverishment of natural vegetation is related to anthropogenic influence in the surrounding region by building a series of models with various biodiversity metrics (community completeness, alpha diversity, beta diversity, gamma diversity, dark diversity and species pool size) as response variables. To estimate the intensity of human activities in the surrounding regions, we used the human footprint index from 2018 (the year our sampling began)—a well-established cumulative metric of human influence[34]—along with all of its eight components, including human population density and various human infrastructure layers. We averaged human influence at various spatial scales around the study region (radii from 10 km to 400 km), because human influence can reach far from mapped features. For example, poaching and logging can occur tens of kilometres from human settlements[35] and are facilitated by many unmapped 'ghost roads' that start from documented roads and lead into natural areas[36]. Similarly, anthropogenic ignition of fires can occur hundreds of kilometres from main roads[37]. Aerial pollution is often deposited several hundreds of kilometres from its source[35], and land use can change local climate over similar scales[38]. To account for the effects of natural processes on biodiversity (for example, geodiversity, habitat patchiness and likelihood of natural fires), we included in our statistical models variables describing climatic, soil and topographic conditions, which we derived from global GIS layers and summarized using four principal component axes. Using fivefold spatial block cross-validation, we determined that linear models produced lower prediction errors with test data, compared with nonlinear alternatives (around 20% versus 40%). We therefore used linear models in further analyses.

The human footprint index and community completeness exhibited a robust negative linear relationship (Fig. 2a), which was already significant when the average human footprint index within a 50-km radius around the site was used, but became even more pronounced when radii of 300 km or larger were considered (Fig. 2b and Extended Data Table 1). In the sampled regions with minimal human footprint index values (close to zero), an average of 35% of suitable species were found in the 100-m^2 sites, but this proportion declined to less than 20% in regions with high human impact. However, there was still variation in community completeness at both the low and the high ends of the human footprint index, showing that sites do not respond uniformly. In contrast to community completeness, alpha diversity was not strongly related to the human footprint index, and nor were the other tested metrics, except beta diversity (Extended Data Fig. 4 and Extended Data Table 1). These results are consistent with our hypothesis that local biodiversity is lower in natural vegetation surrounded by regions with more human activity, but this effect was evident only when we considered community completeness. Raw estimates of alpha diversity were strongly influenced by the wide natural variation in diversity potential determined by the specific biogeographical history of each region. Our results were consistent for six of the eight individual components of the human footprint index: human population density, the extent of electric infrastructure, railways, roads, built environments and croplands all exhibited negative relationships with community completeness (Fig. 2b and Extended Data Table 2). The extent of pasture was an exception to this pattern, because it was not negatively related to community completeness. This could be due to the influence of semi-natural grasslands, in which long-term moderate human influence, including grazing of domestic animals, cultural burning and haymaking, has resulted in highly diverse and well-functioning ecosystems, exemplifying how certain human activities can actually promote native biodiversity[39]. We found that the effect of the human footprint index was strongest when averaged over a range of several hundred kilometres. Besides incorporating far-reaching human influence, larger scales might also more accurately capture cumulative human influence in a particular region over long time periods[40]. However, including in the model a variable representing change in the index between 2000 and 2013 did not reduce the Akaike information criterion (AIC) by more than two units, which suggests that anthropogenic effects have operated over longer timescales. To account for the effects of natural processes

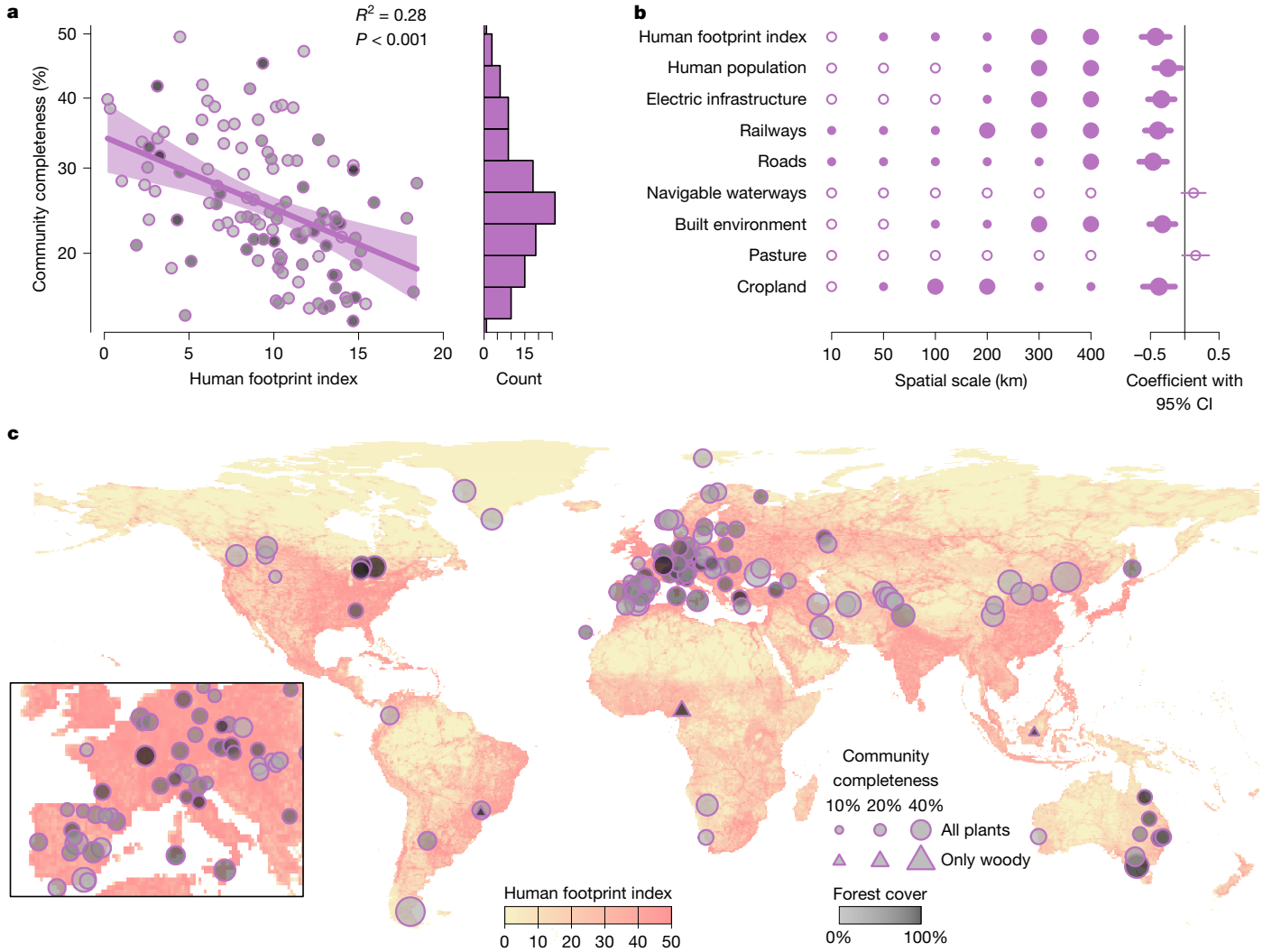


Fig. 2 | Plant diversity in natural vegetation in relation to human effects in the surrounding regions. a, Relationship between community completeness in natural vegetation and the human footprint index in the surrounding area, defined by a radius of 300 km. The prediction line from a multiple linear regression model is shown with the 95% confidence intervals. Note that community completeness values on the *y* axis are back-transformed from the logit scale. The symbol tones indicate forest cover (0–100%). R^2 value of the model and two-tailed *P* value of the relationship are shown; $n = 116$ regions. The distribution of community completeness is shown in the histogram on the right (median, 25%). **b**, Left, model summaries linking community completeness to the human footprint index and its components across spatial scales. Human influence was averaged over various spatial scales around the study regions (radii 10 km, 50 km, 100 km, 200 km, 300 km and 400 km), and the respective models were compared using the Akaike information criterion (AIC). Filled symbols indicate significant relationships ($P < 0.05$), and the large symbol indicates the set of best significant models ($\Delta AIC < 2$). Right, from the best model (the smallest scale at which $\Delta AIC < 2$), the effect of the human footprint index or one of its components is shown as a standardized coefficient (dot) with a 95% confidence interval (CI; line); $n = 116$ regions. Filled symbols and bold confidence interval lines indicate significant effects. **c**, Map of sampling regions, with community completeness indicated by symbol size and the underlying map showing the global variation in the human footprint index[34] (the highest value within each grid cell of around 0.25° × 0.25°). The inset shows part of Europe containing a large number of study regions. Triangles indicate regions in which only woody species were sampled. Symbol tones indicate the percentage of forests in regions.

on community completeness, our models included environmental variables. We found that community completeness decreased along the first principal component (Extended Data Table 1). Thus, suitable species are more likely to fall into the dark diversity in regions characterized by acidic organic soils and higher precipitation (see correlations of principal component axes in Extended Data Fig. 5). Dark diversity, gamma diversity and species pool size increased along the first axis (representing higher soil carbon content, acidity and precipitation; Extended Data Table 1). Alpha and beta diversities showed no significant relationships with the environmental axes.

The negative effect of human activities on community completeness might be associated with several phenomena. Human activities might have led to the fragmentation or reduction of suitable habitats, resulting in smaller populations that are more susceptible to random extinction[9]. In addition, habitat loss is likely to have decreased connectivity between remaining patches of natural vegetation, making it difficult for species to move between areas[41], and defaunation might have disrupted plant seed dispersal networks[30]. Beyond habitat loss, some anthropogenic disturbances, such as tree cutting, illegal harvesting of plants and human-induced wildfires, can cause local extinctions in natural vegetation[10,42]. Moreover, regional human impact can affect natural ecosystems through pollution from roads and other human infrastructure; eutrophication is the most serious threat to plant diversity, because it disproportionally favours a few competitively superior species at the expense of a greater number of other species[8].

Using average human influence as an explanatory variable can mask differences between regions. For example, regions that comprise both highly modified areas (for example, cities) and nature reserves, as well as those experiencing moderate human influence throughout (for example, agricultural landscapes with smaller settlements), might both exhibit an intermediate level of average human influence. We therefore tested how the distribution of the human footprint index within regions affected community completeness. Notably, we found that community completeness had an even stronger negative relationship with anthropogenic influence when we used the 30% quantile of the human footprint index values found within regions (Extended Data Fig. 6). This result suggests that completeness is determined mainly by the extent to which the most natural areas in a region already experience human influence. The idea that 30% coverage of natural vegetation in a landscape supports the persistence of many specialist taxa was proposed previously[43], and aligns with the global target of the Convention on Biological Diversity to protect 30% of land by the year 2030. Our results therefore underscore the importance of devising regional-scale conservation strategies that include maintaining well-preserved natural areas[44].

The turnover of site-specific species pools within regions (structured beta diversity) was significantly positively associated with the human footprint index (Extended Data Fig. 4 and Supplementary Table 1). This might reflect a human preference for naturally diverse regions with a range of different resources[45]. Alternatively, human activities could have promoted plant diversity over millennia by expanding semi-natural habitats and modifying natural ecosystems[39]. Most components of the human footprint index generally exhibited similar relationships, except for the extent of navigable waterways and pastures, which were negatively related to beta diversity (Supplementary Table 1). It is likely that coastal and riverine regions, and those suitable for livestock grazing, naturally exhibit relatively low variation in vegetation types.

The finding that high human footprint index values in a region are associated with low community completeness persisted in several other robustness tests (Supplementary Methods). Statistical interactions between the human footprint index and environmental gradients did not improve the model. Because naturally high beta diversity might decrease community completeness owing to the spatial separation of ecologically similar sites, and because beta diversity was correlated with human influence, we used structural equation modelling to examine the direct and indirect effects of human influence on community completeness. The negative direct effect of the human footprint index on community completeness persisted even if there was an additional negative direct effect of beta diversity. In addition, the effect of the human footprint index on community completeness was consistent across sampling scales (2,500 m^2 or twice as many sites for species co-occurrences), when we excluded alien or very rare species, when regions with only woody species records were included and when we considered the proportion of forest cover in regions. Community completeness was slightly lower in more forested regions. The most parsimonious explanation for this might be a scaling effect—fewer large plant individuals can fit into a fixed area[46]. We also examined the possible effect of geographically uneven sampling by selecting a single study region from each ecoregion (the anthropogenic effect was always negative), adding the European continent as a factor to the model (the negative relationship remained significant) and investigating model residuals (no significant spatial autocorrelation was apparent). Community completeness was slightly lower in Europe than in other regions, which could reflect a cumulative effect of long-term human influence[40].

This global-scale study reveals general patterns, and linkage to specific drivers is based on ecological interpretation rather than experimentation. It is also clear that the human footprint index does not provide a proxy for all potentially important processes, such as the disruption of biotic interaction networks, increasingly frequent climate extremes or the habitat destruction and fragmentation caused by war. The plethora of processes affecting biodiversity certainly contributes to variation around the general trends revealed by our models. The significant relationships we identified apply to the sampled range of the human footprint index, whereas index values outside this range might produce different relationships. In addition, even if the uneven distribution of study regions did not produce an effect in statistical models, the under-representation of several parts of Africa, the Americas and Asia might mean that some human impacts on biodiversity were not well represented. Future work should examine the exact patterns and processes of natural vegetation impoverishment in these undersampled regions.

Our finding of a globally consistent negative relationship between human influence and local plant diversity in relatively natural vegetation is alarming, because plants form the foundation of all terrestrial ecosystems. Reduced community completeness indicates that many species present in the region do not inhabit suitable sites, and this can affect local ecosystem functioning[47]. Although vegetation functioning depends mainly on the traits of co-existing taxa, the presence of a larger proportion of suitable taxa increases the chance that essential functions are represented[48]. We also found that negative human influence was most evident when considered at a scale spanning several hundred kilometres; in other words, biodiversity in natural ecosystems is reduced far beyond human infrastructure. Therefore, conservation actions and land-use planning should consider not only the observed alpha diversity of a site, but also a broader regional context. Ecology has a rich history of conceptual frameworks for biodiversity across scales, such as species–area relationships[49], alpha–beta–gamma diversity[19], community saturation and assembly[33] and the meta-community concept[50]. Building on this collective knowledge, the dark diversity concept offers a species-oriented toolkit for evaluating community patterns and explaining the underlying processes. By allowing the estimation of a site's biodiversity potential (site-specific species pool) and its realization (community completeness), it fosters the comparative study of biodiversity across regions, ecosystem types and taxonomic groups[2]. This improved understanding could help conservation biologists, land managers and policymakers to prevent further losses of biodiversity[51]. Moreover, while site-specific species pools are not depleted, dark diversity offers a narrow window of opportunity for restoration because it indicates which missing species are still regionally present[52–54].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-08814-5.

1. Ceballos, G. et al. Accelerated modern human-induced species losses: entering the sixth mass extinction. *Sci. Adv.* **1**, e1400253 (2015).
2. Pärtel, M., Szava-Kovats, R. & Zobel, M. Dark diversity: shedding light on absent species. *Trends Ecol. Evol.* **26**, 124–128 (2011).
3. Jaureguiberry, P. et al. The direct drivers of recent global anthropogenic biodiversity loss. *Sci. Adv.* **8**, eabm9982 (2022).
4. Newbold, T. et al. Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015).
5. Williams, N. S. G. et al. A conceptual framework for predicting the effects of urban environments on floras. *J. Ecol.* **97**, 4–9 (2009).
6. Le Provost, G. et al. Grassland-to-crop conversion in agricultural landscapes has lasting impact on the trait diversity of bees. *Landsc. Ecol.* **36**, 281–295 (2021).
7. Haddad, N. M. et al. Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **1**, e1500052 (2015).
8. Stevens, C. J., Thompson, K., Grime, J. P., Long, C. J. & Gowing, D. J. G. Contribution of acidification and eutrophication to declines in species richness of calcifuge grasslands along a gradient of atmospheric nitrogen deposition. *Funct. Ecol.* **24**, 478–484 (2010).

9. Chase, J. M., Blowes, S. A., Knight, T. M., Gerstner, K. & May, F. Ecosystem decay exacerbates biodiversity loss with habitat loss. *Nature* **584**, 238–243 (2020).
10. Barlow, J. et al. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature* **535**, 144–147 (2016).
11. Gray, C. L. et al. Local biodiversity is higher inside than outside terrestrial protected areas worldwide. *Nat. Commun.* **7**, 12306 (2016).
12. Santangeli, A. et al. Mixed effects of a national protected area network on terrestrial and freshwater biodiversity. *Nat. Commun.* **14**, 5426 (2023).
13. Jones, K. R. et al. One-third of global protected land is under intense human pressure. *Science* **360**, 788–791 (2018).
14. Cai, L. et al. Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytol.* **237**, 1432–1445 (2022).
15. Sabatini, F. M. et al. Global patterns of vascular plant alpha diversity. *Nat. Commun.* **13**, 4683 (2022).
16. Fraser, L. H. et al. Coordinated distributed experiments: an emerging tool for testing global hypotheses in ecology and environmental science. *Front. Ecol. Environ.* **11**, 147–155 (2013).
17. Pärtel, M., Szava-Kovats, R. & Zobel, M. Community completeness: linking local and dark diversity within the species pool concept. *Folia Geobot.* **48**, 307–317 (2013).
18. Carmona, C. P. & Pärtel, M. Estimating probabilistic site-specific species pools and dark diversity from co-occurrence data. *Glob. Ecol. Biogeogr.* **30**, 316–326 (2021).
19. Whittaker, R. H. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* **30**, 279–338 (1960).
20. Harrison, S., Vellend, M. & Damschen, E. I. 'Structured' beta diversity increases with climatic productivity in a classic dataset. *Ecosphere* **2**, 1–13 (2011).
21. Anderson, M. J. et al. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecol. Lett.* **14**, 19–28 (2011).
22. Vellend, M. Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).
23. Salguero-Gómez, R. et al. Fast–slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proc. Natl Acad. Sci. USA* **113**, 230–235 (2016).
24. Peco, B., Laffan, S. W. & Moles, A. T. Global patterns in post-dispersal seed removal by invertebrates and vertebrates. *PLoS One* **9**, e91256 (2014).
25. Riibak, K. et al. Drivers of plant community completeness differ at regional and landscape scales. *Agric. Ecosyst. Environ.* **301**, 107004 (2020).
26. Ben-Hur, E. & Kadmon, R. Heterogeneity–diversity relationships in sessile organisms: a unified framework. *Ecol. Lett.* **23**, 193–207 (2019).
27. González-Trujillo, J. D., Román-Cuesta, R. M., Muñiz-Castillo, A. I., Amaral, C. H. & Araújo, M. B. Multiple dimensions of extreme weather events and their impacts on biodiversity. *Clim. Change* **176**, 155 (2023).
28. Gallagher, R. V. et al. High fire frequency and the impact of the 2019–2020 megafires on Australian plant diversity. *Divers. Distrib.* **27**, 1166–1179 (2021).
29. Outhwaite, C. L., McCann, P. & Newbold, T. Agriculture and climate change are reshaping insect biodiversity worldwide. *Nature* **605**, 97–102 (2022).
30. Dirzo, R. et al. Defaunation in the Anthropocene. *Science* **345**, 401–406 (2014).
31. Pärtel, M. et al. DarkDivNet—a global research collaboration to explore the dark diversity of plant communities. *J. Veg. Sci.* **30**, 1039–1043 (2019).
32. Tikhonov, G. et al. Joint species distribution modelling with the R-package HMSC. *Methods Ecol. Evol.* **11**, 442–447 (2020).
33. Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
34. Gassert, F. et al. An operational approach to near real time global high resolution mapping of the terrestrial Human Footprint. *Front. Remote Sens.* **4**, 1130896 (2023).
35. McDonald, R. I. et al. Urban effects, distance, and protected areas in an urbanizing world. *Landsc. Urban Plann.* **93**, 63–75 (2009).
36. Engert, J. E. et al. Ghost roads and the destruction of Asia-Pacific tropical forests. *Nature* **629**, 370–375 (2024).
37. Fusco, E. J., Abatzoglou, J. T., Balch, J. K., Finn, J. T. & Bradley, B. A. Quantifying the human influence on fire ignition across the western USA. *Ecol. Appl.* **26**, 2390–2401 (2016).
38. Lawton, R. O., Nair, U. S., Pielke, R. A. & Welch, R. M. Climatic impact of tropical lowland deforestation on nearby montane cloud forests. *Science* **294**, 584–587 (2001).
39. Bengtsson, J. et al. Grasslands—more important for ecosystem services than you might think. *Ecosphere* **10**, e02582 (2019).
40. Ellis, E. C. Land use and ecological change: a 12,000-year history. *Annu. Rev. Environ. Resour.* **46**, 1–33 (2021).
41. Taylor, P. D., Fahrig, L., Henein, K. & Merriam, G. Connectivity is a vital element of landscape structure. *Oikos* **68**, 571–573 (1993).
42. van Wees, D. et al. The role of fire in global forest loss dynamics. *Glob. Chang. Biol.* **27**, 2377–2391 (2021).
43. Hanski, I. Habitat loss, the dynamics of biodiversity, and a perspective on conservation. *Ambio* **40**, 248–255 (2011).
44. Aavik, T. & Helm, A. Restoration of plant species and genetic diversity depends on landscape-scale dispersal. *Restor. Ecol.* **26**, S92–S102 (2018).
45. Araujo, M. B. The coincidence of people and biodiversity in Europe. *Glob. Ecol. Biogeogr.* **12**, 5–12 (2003).
46. Oksanen, J. Is the humped relationship between species richness and biomass an artefact due to plot size? *J. Ecol.* **84**, 293–295 (1996).
47. Hooper, D. U. et al. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
48. Reich, P. B. et al. Impacts of biodiversity loss escalate through time as redundancy fades. *Science* **336**, 589–592 (2012).
49. Storch, D. The theory of the nested species–area relationship: geometric foundations of biodiversity scaling. *J. Veg. Sci.* **27**, 880–891 (2016).
50. Leibold, M. A. et al. The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* **7**, 601–613 (2004).
51. Leclère, D. et al. Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature* **585**, 551–556 (2020).
52. Lewis, R. J. et al. Applying the dark diversity concept to nature conservation. *Conserv. Biol.* **31**, 40–47 (2017).
53. Deschênes, É., Santala, K. R., Lavigne, J. & Aubin, I. Using a trait-based dark diversity approach to evaluate natural recovery potential in forests. *Restor. Ecol.* **32**, e14251 (2024).
54. Moeslund, J. E. et al. Using dark diversity and plant characteristics to guide conservation and restoration. *J. Appl. Ecol.* **54**, 1730–1741 (2017).

Meelis Pärtel[1]✉, Riin Tamme[1], Carlos P. Carmona[1], Kersti Riibak[1], Mari Moora[1], Jonathan A. Bennett[2], Alessandro Chiarucci[3], Milan Chytrý[4], Francesco de Bello[5,6], Ove Eriksson[7], Susan Harrison[8], Robert John Lewis[9], Angela T. Moles[10], Maarja Öpik[1], Jodi N. Price[11], Vistorina Amputu[12], Diana Askarizadeh[13,14], Zohreh Atashgahi[15], Isabelle Aubin[16], Francisco M. Azcárate[17,18], Matthew D. Barrett[19], Maral Bashirzadeh[20], Zoltán Bátori[21], Natalie Beenaerts[22], Kolja Bergholz[23], Kristine Birkeli[24,25], Idoia Biurrun[26], José M. Blanco-Moreno[27,28], Kathryn J. Bloodworth[29], Laura Boisvert-Marsh[16], Bazartseren Boldgiv[30], Pedro H. S. Brancalion[31,32], Francis Q. Brearley[33], Charlotte Brown[34,35], C. Guillermo Bueno[36], Gabriella Buffa[37], James F. Cahill[35], Juan A. Campos[26], Giacomo Cangelmi[38], Michele Carbognani[39], Christopher Carcaillet[40,41], Bruno E. L. Cerabolini[42], Richard Chevalier[43], Jan S. Clavel[44], José M. Costa[45], Sara A. O. Cousins[46], Jan Čuda[47], Mariana Dairel[48], Michele Dalle Fratte[42], Alena Danilova[49], John Davison[1], Balázs Deák[50], Silvia Del Vecchio[3], Iwona Dembicz[51], Jürgen Dengler[52], Jiri Dolezal[6,53], Xavier Domene[54,55], Miroslav Dvorsky[53], Hamid Ejtehadi[56], Lucas Enrico[57,58], Dmitrii Epikhin[59], Anu Eskelinen[60,61], Franz Essl[62], Gaohua Fan[63], Edy Fantinato[37], Fatih Fazlioglu[64,65], Eduardo Fernández-Pascual[66,67], Arianna Ferrara[3], Alessandra Fidelis[48], Markus Fischer[68], Maren Flagmeier[69], T'ai G. W. Forte[39], Lauchlan H. Fraser[70], Junichi Fujinuma[1], Fernando F. Furquim[71], Berle Garris[72], Heath W. Garris[73], Melisa A. Giorgis[57,58], Gianpietro Giusso del Galdo[74], Ana González-Robles[75,76], Megan K. Good[77], Moisès Guardiola[78], Riccardo Guarino[79], Irene Guerrero[17], Joannès Guillemot[31,80,81], Behlül Güler[82], Yinjie Guo[83], Stef Haesen[84,85], Martin Hejda[47], Ruben H. Heleno[45], Toke T. Høye[86,87], Richard Hrivnák[88], Yingxin Huang[89,90,91], John T. Hunter[92], Dmytro Iakushenko[93,94], Ricardo Ibáñez[95], Nele Ingerpuu[1], Severin D. H. Irl[96], Eva Janíková[6], Florian Jansen[97], Florian Jeltsch[23], Anke Jentsch[98], Borja Jiménez-Alfaro[66,67], Madli Jõks[1], Mohammad H. Jouri[99], Sahar Karami[56], Negin Katal[100], András Kelemen[21,50], Bulat I. Khairullin[101], Anzar A. Khuroo[102], Kimberly J. Komatsu[29], Marie Konečná[6], Ene Kook[1], Lotte Korell[61,103], Natalia Koroleva[49], Kirill A. Korznikov[104], Maria V. Kozhevnikova[101], Łukasz Kozub[51], Lauri Laanisto[105], Helena Lager[106], Vojtech Lanta[104], Romina G. Lasagno[107], Jonas J. Lembrechts[44,108], Liping Li[109], Aleš Lisner[6], Houjia Liu[63], Kun Liu[110], Xuhe Liu[89,90,91,111], Manuel Esteban Lucas-Borja[112], Kristin Ludewig[113], Katalin Lukács[50], Jona Luther-Mosebach[114], Petr Macek[105,115], Michela Marignani[116], Richard Michalet[117], Tamás Miglécz[118], Jesper Erenskjold Moeslund[86], Karlien Moeys[84], Daniel Montesinos[45,119,120], Eduardo Moreno-Jiménez[69,121], Ivan Moysiyenko[122], Ladislav Mucina[123,124], Miriam Muñoz-Rojas[125,126], Raytha A. Murillo[35], Sylvia M. Nambahu[127], Lena Neuenkamp[68,128], Signe Normand[129], Arkadiusz Nowak[130], Paloma Nuche[131], Tatjana Oja[1], Vladimir G. Onipchenko[132], Kalina L. Pachedjieva[133], Bruno Paganeli[1], Begoña Peco[17], Ana M. L. Peralta[134], Aaron Pérez-Haase[27,28], Pablo L. Peri[107,135], Alessandro Petraglia[39], Gwendolyn Peyre[136], Pedro Antonio Plaza-Álvarez[112], Jan Plue[137], Honor C. Prentice[138], Vadim E. Prokhorov[101], Dajana Radujković[44], Soroor Rahmanian[61,139], Triin Reitalu[1,140], Michael Ristow[23], Agnès A. Robin[80,81,141], Ana Belén Robles[142], Daniel A. Rodríguez Ginart[5], Raúl Román[143], Ruben E. Roos[144,145], Leonardo Rosati[146], Jiří Sádlo[47], Karina Salimbayeva[35], Rut Sánchez de Dios[147], Khaliun Sanchir[30], Cornelia Sattler[148], John D. Scasta[149], Ute Schmiedel[150], Julian Schrader[148], Nick L. Schultz[151], Giacomo Sellan[152], Josep M. Serra-Diaz[153], Giulia Silan[37], Hana Skálová[47], Nadiia Skobel[51,122], Judit Sonkoly[154], Kateřina Štajerová[47], Ivana Svitková[88], Sebastian Świerszcz[130,155], Andrew J. Tanentzap[156], Fallon M. Tanentzap[157], Rubén Tarifa[75,158], Pablo Tejero[36], Dzhamal K. Tekeev[159], Michael Tholin[106], Ruben S. Thormodsæter[24], Yichen Tian[109], Alla Tokaryuk[160], Csaba Tölgyesi[21], Marcello Tomaselli[39], Enrico Tordoni[1], Péter Török[161], Béla Tóthmérész[162], Aurèle Toussaint[163], Blaise Touzard[117], Diego P. F. Trindade[1,5], James L. Tsakalos[123,164], Sevda Türkiş[165], Enrique Valencia[147], Mercedes Valerio[6,95], Orsolya Valkó[50], Koenraad Van Meerbeek[84,85], Vigdis Vandvik[24,25], Jesus Villellas[166], Risto Virtanen[60], Michaela Vítková[47], Martin Vojík[47,167,168], Andreas von Hessberg[98], Jonathan von Oppen[169,170], Viktoria Wagner[35], Ji-Zhong Wan[171], Chun-Jing Wang[172], Sajad A. Wani[102], Lina Weiss[23,173], Tricia Wevill[174], Sa Xiao[110], Oscar Zárate Martínez[1] & Martin Zobel[1]

[1]Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia. [2]Department of Plant Sciences, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. [3]Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum —University of Bologna, Bologna, Italy. [4]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic. [5]CIDE, CSIC-UV-GVA, Valencia, Spain. [6]Department of Botany,

Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. [7]Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden. [8]Department of Environmental Science and Policy, University of California Davis, Davis, CA, USA. [9]Norwegian Institute for Nature Research, Bergen, Norway. [10]Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, UNSW Sydney, Sydney, New South Wales, Australia. [11]Gulbali Institute, Charles Sturt University, Albury, New South Wales, Australia. [12]Plant Ecology Group, Institute of Evolution and Ecology, University of Tübingen, Tübingen, Germany. [13]Independent researcher, Tehran, Iran. [14]Department of Reclamation of Arid and Mountainous Regions, University of Tehran, Tehran, Iran. [15]Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran. [16]Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, Sault Ste Marie, Ontario, Canada. [17]Terrestrial Ecology Group, Department of Ecology, Universidad Autónoma de Madrid, Madrid, Spain. [18]Centro de Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, Spain. [19]Australian Tropical Herbarium, James Cook University, Smithfield, Queensland, Australia. [20]Department of Range and Watershed Management, Faculty of Natural Resources and Environment, Ferdowsi University of Mashhad, Mashhad, Iran. [21]Department of Ecology, University of Szeged, Szeged, Hungary. [22]Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium. [23]Plant Ecology and Nature Conservation, University of Potsdam, Potsdam, Germany. [24]Department of Biological Sciences, University of Bergen, Bergen, Norway. [25]Bjerknes Centre for Climate Research, University of Bergen, Bergen, Norway. [26]Department of Plant Biology and Ecology, University of the Basque Country UPV/EHU, Bilbao, Spain. [27]Department of Evolutionary Biology, Ecology and Environmental Sciences (Botany and Mycology), Universitat de Barcelona, Barcelona, Spain. [28]Biodiversity Research Institute (IRBio), Universitat de Barcelona, Barcelona, Spain. [29]Department of Biology, University of North Carolina at Greensboro, Greensboro, NC, USA. [30]Department of Biology, National University of Mongolia, Ulaanbaatar, Mongolia. [31]Department of Forest Sciences, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, Brazil. [32]Re.green, Rio de Janeiro, Brazil. [33]Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK. [34]Département de biologie, Université de Sherbrooke, Sherbrooke, Quebec, Canada. [35]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada. [36]Instituto Pirenaico de Ecologia, CSIC, Jaca, Spain. [37]Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy. [38]Department of Life, Health and Environmental Science, University of L'Aquila, Coppito, L'Aquila, Italy. [39]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy. [40]École Pratique des Hautes Études, Paris Sciences Lettres University (EPHE-PSL), Paris, France. [41]University Claude Bernard Lyon 1, LEHNA UMR5023, CNRS, ENTPE, Villeurbanne, France. [42]Department of Biotechnology and Life Science, University of Insubria, Varese, Italy. [43]Conservatoire d'espaces naturels Centre–Val de Loire, Orléans, France. [44]Research Group Plants and Ecosystems (PLECO), University of Antwerp, Wilrijk, Belgium. [45]Centre for Functional Ecology, Associate Laboratory TERRA, Department of Life Sciences, University of Coimbra, Coimbra, Portugal. [46]Department of Physical Geography, Stockholm University, Stockholm, Sweden. [47]Department of Invasion Ecology, Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic. [48]Instituto de Biociências, Lab of Vegetation Ecology, Universidade Estadual Paulista (UNESP), Rio Claro, Brazil. [49]Independent researcher, Kirovsk, Russia. [50]Lendület Seed Ecology Research Group, Institute of Ecology and Botany, HUN-REN Centre for Ecological Research, Vácrátót, Hungary. [51]Institute of Environmental Biology, Faculty of Biology, University of Warsaw, Warsaw, Poland. [52]Vegetation Ecology Research Group, Institute of Natural Resource Sciences (IUNR), Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland. [53]Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic. [54]CREAF (Centre for Ecological Research and Forestry Applications), Bellaterra, Spain. [55]Universitat Autònoma de Barcelona, Bellaterra, Spain. [56]Quantitative Plant Ecology and Biodiversity Research Lab, Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashdad, Iran. [57]Instituto Multidisciplinario de Biología Vegetal (CONICET-UNC), Córdoba, Argentina. [58]FCEFyN, Universidad Nacional de Córdoba, Córdoba, Argentina. [59]Independent researcher, Moscow, Russia. [60]Department of Ecology and Genetics, University of Oulu, Oulu, Finland. [61]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [62]Division of BioInvasions, Global Change and Macroecology, University of Vienna, Vienna, Austria. [63]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China. [64]Chair of Plant Ecology, University of Bayreuth, Bayreuth, Germany. [65]Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Ordu University, Ordu, Turkey. [66]Biodiversity Research Institute (IMIB), University of Oviedo–CSIC–Principality of Asturias, Mieres, Spain. [67]Department of Organismal and Systems Biology, University of Ovidedo, Oviedo, Spain. [68]Institute of Plant Sciences, University of Bern, Bern, Switzerland. [69]Department of Agricultural and Food Chemistry, Universidad Autónoma de Madrid, Madrid, Spain. [70]Department of Natural Resource Sciences, Thompson Rivers University, Kamloops, British Columbia, Canada. [71]Graduate Program in Botany, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [72]Independent researcher, Mancelona, MI, USA. [73]Au Sable Institute of Environmental Studies, Mancelona, MI, USA. [74]Department of Geological, Biological and Environmental Sciences, University of Catania, Catania, Italy. [75]Departamento de Biología Animal, Biología Vegetal y Ecología, Universidad de Jaén, Jaén, Spain. [76]Instituto Interuniversitario del Sistema Tierra de Andalucía, Universidad de Jaén, Jaén, Spain. [77]School of Agriculture, Food and Ecosystem Sciences, University of Melbourne, Melbourne, Victoria, Australia. [78]Unit of Botany, Department of Animal and Plant Biology and Ecology, Universitat Autònoma de Barcelona, Bellaterra, Spain. [79]Department of Biological, Chemical and Pharmaceutical Sciences and Technologies, University of Palermo, Palermo, Italy. [80]CIRAD, UMR Eco&Sols, Montpellier, France. [81]Eco&Sols, University Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France. [82]Biology Education, Dokuz Eylül University, Buca, Turkey. [83]Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China. [84]Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium. [85]KU Leuven Plant Institute, KU Leuven, Leuven, Belgium. [86]Department of Ecoscience, Aarhus University, Aarhus, Denmark. [87]Arctic Research Centre, Aarhus University, Aarhus, Denmark. [88]Institute of Botany, Plant Science and Biodiversity Center, Slovak Academy of Sciences, Bratislava, Slovakia. [89]State Key Laboratory of Black Soils Conservation and Utilization, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China. [90]Jilin Songnen Grassland Ecosystem National Observation and Research Station, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China. [91]Jilin Provincial Key Laboratory of Grassland Farming, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China. [92]School of Environmental and Rural Science, University of New England, Armidale, New South Wales, Australia. [93]Institute of Biological Sciences, University of Zielona Góra, Zielona Góra, Poland. [94]F. Falz-Fein Biosphere Reserve Askania Nova, Kyiv, Ukraine. [95]Departamento de Biología Ambiental, Facultad de Ciencias, Universidad de Navarra, Pamplona, Spain. [96]Biogeography and Biodiversity Lab, Institute of Physical Geography, Goethe-University Frankfurt, Frankfurt am Main, Germany. [97]Faculty of Agricultural and Environmental Sciences, University of Rostock, Rostock, Germany. [98]Department of Disturbance Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany. [99]Department of Range and Watershed Management, Faculty of Natural Resources, Islamic Azad University Nour Branch, Nour, Iran. [100]Chair of Sensor-based Geoinformatics, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany. [101]Independent researcher, Kazan, Russia. [102]Centre for Biodiversity and Taxonomy, Department of Botany, University of Kashmir, Srinagar, India. [103]Department of Species Interaction Ecology, Helmholtz Centre for Environmental Research—UFZ, Leipzig, Germany. [104]Department of Functional Ecology, Institute of Botany, Czech Academy of Sciences, Třeboň, Czech Republic. [105]Chair of Biodiversity and Nature Tourism, Estonian University of Life Sciences, Tartu, Estonia. [106]Kalmar County Administrative Board, Färjestaden, Sweden. [107]Instituto Nacional de Tecnología Agropecuaria (INTA), Río Gallegos, Argentina. [108]Ecology and Biodiversity (E&B), Utrecht University, Utrecht, The Netherlands. [109]Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. [110]State Key Laboratory of Grassland and Agro-Ecosystems, School of Life Sciences, Lanzhou University, Lanzhou, China. [111]School of Grassland Science, Beijing Forestry University, Beijing, China. [112]Higher Technical School of Agricultural and Forestry Engineering, Castilla-La Mancha University, Albacete, Spain. [113]Applied Plant Ecology, Institute of Plant Science and Microbiology, University of Hamburg, Hamburg, Germany. [114]Netzwerk für Angewandte Ökologie, Hamburg, Germany. [115]Institute of Hydrobiology, Biology Centre of the Czech Academy of Sciences, Ceske Budejovice, Czech Republic. [116]Department of Life and Environmental Sciences, University of Cagliari, Cagliari, Italy. [117]Univ. Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, Pessac, France. [118]ÖMKi—Research Institute of Organic Agriculture, Budapest, Hungary. [119]Australian Tropical Herbarium, James Cook University, Cairns, Queensland, Australia. [120]College of Science and Engineering, James Cook University, Cairns, Queensland, Australia. [121]Institute for Advanced Research in Chemical Sciences (IAdChem), Universidad Autónoma de Madrid, Madrid, Spain. [122]Department of Botany, Kherson State University, Kherson, Ukraine. [123]Harry Butler Institute, Murdoch University, Perth, Western Australia, Australia. [124]Department of Geography and Environmental Studies, Stellenbosch University, Stellenbosch, South Africa. [125]Laboratorio de Biodiversidad y Funcionamiento Ecosistémico Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, Sevilla, Spain. [126]Centre for Ecosystem Science, UNSW Sydney, Sydney, New South Wales, Australia. [127]Department of Agriculture and Natural Resource Sciences, Namibia University of Science and Technology, Windhoek, Namibia. [128]Institute of Landscape Ecology, University of Münster, Münster, Germany. [129]Center for Sustainable Landscapes Under Global Change, Department of Biology, Aarhus University, Aarhus, Denmark. [130]Botanical Garden, Center for Biological Diversity Conservation, Polish Academy of Sciences, Warszawa, Poland. [131]Greenpeace España, Madrid, Spain. [132]Shenzhen MSU-BIT University, Shenzhen, China. [133]Department of Ecology and Environmental Protection, Faculty of Biology, Sofia University St Kliment Ohridski, Sofia, Bulgaria. [134]Departemento Biología y Geología, Física y Química Inorgánica, Universidad Rey Juan Carlos, Móstoles, Spain. [135]Universidad Nacional de la Patagonia Austral (UNPA), CONICET, Río Gallegos, Argentina. [136]Department of Civil and Environmental Engineering, University of the Andes, Bogotá, Colombia. [137]Swedish Biodiversity Centre, Department of Urban and Rural Development, Swedish University of Agricultural Sciences, Uppsala, Sweden. [138]Department of Biology, Lund University, Lund, Sweden. [139]Remote Sensing Centre for Earth System Research (RSC4Earth), Leipzig University, Leipzig, Germany. [140]Institute of Geology, Tallinn University of Technology, Tallinn, Estonia. [141]Department of Soil Sciences, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, Brazil. [142]Estación Experimental del Zaidín (CSIC), Granada, Spain. [143]Department of Agronomy, University of Almería, Almería, Spain. [144]Norwegian Institute for Nature Research, Oslo, Norway. [145]Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway. [146]Scuola di Scienze Agrarie, Forestali, Alimentari e Ambientali, Università della Basilicata, Potenza, Italy. [147]Departamento de Biodiversidad, Ecología y Evolución, Facultad de Ciencias Biológicas, Universidad Complutense de Madrid, Madrid, Spain. [148]School of Natural Sciences, Macquarie University, Sydney, New South Wales, Australia. [149]Department of Ecosystem Science and Management, Laramie Research and Extension Center, University of Wyoming, Laramie, WY, USA. [150]Institute of Plant Science and Microbiology, University of Hamburg, Hamburg, Germany. [151]Future Regions Research Centre, Federation University Australia, Ballarat, Victoria, Australia. [152]CIRAD-UMR EcoFoG, Kourou, French Guiana. [153]Botanical Institute of Barcelona (CSIC-CMCNB), Barcelona, Spain. [154]Department of Ecology, University of Debrecen, Debrecen, Hungary. [155]Institute of Agroecology and Plant Production, Wrocław University of Environmental and Life Sciences, Wrocław, Poland. [156]Ecosystems and Global Change Group, School of the Environment, Trent University, Peterborough, Ontario, Canada. [157]Department of Plant Sciences, University of Cambridge, Cambridge, UK. [158]Estación Experimental de Zonas Áridas (EEZA-CSIC), Almería, Spain. [159]Independent reseacher, Teberda, Russia. [160]Yuriy Fedkovych Chernivtsi National University,

Chernivtsi, Ukraine. [161]HUN-REN-UD Functional and Restoration Ecology Research Group, Department of Ecology, University of Debrecen, Debrecen, Hungary. [162]HUN-REN-UD Biodiversity and Ecosystem Services Research Group, Department of Ecology, University of Debrecen, Debrecen, Hungary. [163]Centre de Recherche sur la Biodiversité et l'Environnement (CRBE), UMR 5300 UPS-CNRS-IRD-INP, Université Paul Sabatier–Toulouse 3, Toulouse, France. [164]School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy. [165]Department of Mathematics and Science Education, Faculty of Education, Ordu University, Ordu, Turkey. [166]Departamento de Ciencias de la Vida, Universidad de Alcalá, Alcalá de Henares, Spain. [167]Department of Applied Ecology, Faculty of Environmental Sciences, Czech University of Life Sciences, Prague, Czech Republic. [168]Nature Conservation Agency of the Czech Republic, Prague, Czech Republic. [169]Department of Biology, Aarhus University, Aarhus, Denmark. [170]Department of Environmental Sciences, University of Basel, Basel, Switzerland. [171]Key Laboratory of Mountain Surface Processes and Ecological Regulation, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, China. [172]Sichuan Academy of Forestry, Chengdu, China. [173]National Monitoring Centre for Biodiversity Germany, Leipzig, Germany. [174]Deakin University, Burwood, Victoria, Australia. ✉e-mail: meelis.partel@ut.ee

Methods

DarkDivNet sampling scheme

In 2018, we launched a global collaborative research consortium to sample both locally observed alpha diversity and dark diversity of terrestrial plant communities using a standardized methodology. A detailed sampling protocol was produced before fieldwork began[31]. Each study region covered an area of approximately 300 km^2, defined by a circle of 20-km diameter with the available area influenced by geographical and practical limitations (coastline, private ownership and other access restrictions). This spatial scale was selected on the basis of the authors' expertise, in the expectation that it would incorporate areas with a relatively uniform biogeographical history while still exhibiting variation in natural vegetation. In addition, mechanisms of long-distance seed dispersal can operate at this scale[55]. In each region, we defined at least 30 sites, in which we sampled a 100-m^2 (10 m × 10 m) area by recording all vascular plant species. Where feasible, we sampled more sites in the region to examine how sampling intensity might affect the results. The sites were selected to proportionally represent the typical natural vegetation types of the region without major human influence. These included semi-natural grasslands, representing habitats that have developed over thousands of years through grazing by domestic animals and mowing, and forests that had been managed with low intensity and had species composition and tree-layer structure similar to old-growth forests. Here we report the results from 5,415 sites in 119 regions for which sampling was completed by 1 February 2024 (Supplementary Table 2).

To assess whether dark diversity methods could predict species that were absent from the 100-m^2 area but present in its immediate vicinity, and to estimate the effect of spatial scale on dark diversity, we selected one to three sites per region in which we sampled vascular plants in a 2,500-m^2 (50 m × 50 m) area within which the 100-m^2 area was nested. In four regions, sampling of the larger area was not possible or the large area had no new taxa, so these regions were omitted from the respective test. In addition, in 76 regions, we had sufficient expertise to assess which of the species found in the region were ecologically well-suited for a selected site (that is, belonging to the site-specific species pool). This information allowed us to test the applicability of dark diversity methods within our sampling framework (see below).

Biodiversity metrics

Biodiversity metrics were determined for each site in each region (Fig. 1). Alpha diversity A was defined as the number of vascular plant species found in the 100-m^2 area describing a site (Fig. 1a). Dark diversity D was quantified for each site k by examining species co-occurrences within the surrounding region using the hypergeometric method, implemented in the R package DarkDiv (ref. 18). This technique uses information about how each species i that is absent from the study site but present in the surrounding region co-occurs with species j that is present at the study site. If an absent species co-occurs more frequently with observed species than it would do under random expectation, it is likely to belong to the dark diversity. The expected number of co-occurrences is mathematically defined by the hypergeometric distribution. For each pair of absent and present species, we compared the observed number of co-occurrences M_{ij} with the expected value, which is defined as the mean of the hypergeometric distribution:

$$\overline{M}_{ij} = \frac{n_i n_j}{N}$$

where n_i and n_j are the total number of occurrences of species i and j, respectively, and N is the total number of sites sampled in that region. The standardized effect size (SES) was used as an indicator of the suitability of absent species i for site k on the basis of co-occurrences with present species j (Fig. 1b), and was calculated as the difference between the observed and the expected numbers of co-occurrences divided by the standard deviation of the expected number of co-occurrences, as derived from the hypergeometric distribution:

$$\mathrm{SES}_{ij} = \frac{M_{ij} - \overline{M}_{ij}}{\sqrt{\left(\frac{n_i n_j}{N}\right)\left(\frac{N-n_i}{N}\right)\left(\frac{N-n_j}{N-1}\right)}}$$

We estimated the suitability of site k for all species i absent from the site but present in the region, by averaging suitability indicator values from all present species j using the number of species found in site k (n_k):

$$\mathrm{SES}_{ki} = \frac{\sum_j^{n_k} \mathrm{SES}_{ij}}{n_k}$$

The SES_{ki} values were subsequently transformed to a 0–1 scale by applying inverse probit transformation, which places the SES_{ki} value within the cumulative normal distribution function with mean = 0 and standard deviation = 1 (Fig. 1c):

$$P_{ki} = \int_{-\infty}^{\mathrm{SES}_{ki}} \frac{e^{-\left(\frac{\mathrm{SES}_{ki}^2}{2}\right)}}{\sqrt{2\pi}}$$

This estimate expressed the probability that species i belonged to the dark diversity of site k. Our estimated dark diversity probabilities were supported by two independent tests, one investigating which absent species were found in the immediate vicinity of a site and another using expert assessment (Extended Data Fig. 2). We also considered how the suitability of absent species might be estimated using an alternative technique—joint species distribution modelling (JSDM)[56] (Supplementary Methods).

Dark diversity size for a study site was the sum of the probabilities P_{ki} of all locally absent species found elsewhere in the region (Fig. 1d). For co-occurrences, we always considered 30 sites (each described by a 100-m^2 area) within the same region (Fig. 1a), which is the minimum number sampled and generally sufficient for the method[18]. For regions with more than 30 sampled sites, we used an iterative procedure, each time randomly selecting 30 sites for species co-occurrences. Dark diversity size in those regions was estimated as the median from 100 iterations. Similarly, estimates of gamma diversity G were obtained using iteration, taking the median cumulative species number from 30 sites in a region. To test whether 30 sites was sufficient to estimate the variation in regional richness, we estimated species richness with complete sample coverage using incidence-based extrapolation based on the Bernoulli product model[57], implemented within the R iNEXT package[58]. Gamma diversity from 30 sites correlated strongly with the extrapolated value (Spearman r = 0.95; Extended Data Fig. 3a)

Using alpha, dark and gamma diversities for each site, we calculated: species pool size as the sum of alpha and dark diversity: $P = A + D$; community completeness as the percentage of alpha diversity among all suitable species for that site: $C = A/(A + D) \times 100\%$; and beta diversity as the percentage of gamma diversity belonging to other species pools in the region and unsuitable for the specific site: $B = (G - A - D)/G \times 100\%$ (Fig. 1d). This metric is identical to Whittaker's effective turnover at the species pool level, expressed as a percentage rather than a ratio (G/P) − 1. In analyses, all biodiversity metrics were transformed to improve distributions: those based on counts or sums (alpha, dark and gamma diversity, species pool size) were log-transformed, and those based on percentages (community completeness and beta diversity) were logit transformed. To aid intuitive understanding, we show untransformed values on graph axes. Because several of the diversity metrics are either subsets or calculated from each other, it is expected that these are closely related. However, bivariate relationships between our study variables (Extended Data Fig. 7) showed that all metrics bear

some independent information, and the variability among and within regions is large.

All of our biodiversity metrics depend on the sampling scheme, including characteristics such as sample area or number of sites. To investigate how much our biodiversity metrics change if using a larger sample area, we used 1–3 sites in each region where both 100-m^2 and 2,500-m^2 areas were sampled. Similarly, we examined the effect of using a larger number of sites to characterize co-occurrences; using 60 sites from 27 regions where they were available. Overall, global variation in our metrics was highly correlated regardless of sample area and the number of sites considered (Spearman correlation > 0.8; Extended Data Fig. 3b–k).

According to the DarkDivNet protocol, in very diverse tropical regions we only sampled woody vascular plant species. Although alpha diversity, dark diversity, species pool size and gamma diversity are evidently smaller when herbaceous species are omitted, community completeness and beta diversity should still be relatively comparable with other regions because these metrics are unitless. To ensure full comparability between biodiversity metrics, we used only the 116 regions in which all vascular plants were sampled in the main analyses, but repeated the main tests for community completeness with all 119 regions within robustness analyses (Supplementary Methods).

Statistical details on the contribution to variation in alpha diversity

Alpha diversity can be seen as a subset of gamma diversity in which the species pool has been filtered according to beta diversity, and the realization of the species pool is defined by community completeness (Fig. 1). We examined how much of the variation in alpha diversity is determined by variation in gamma diversity, beta diversity (these two define the site-specific species pool size) and community completeness. We randomly selected one site from each region in order to have independent local and regional variables (gamma diversity is the same for all sites in a region). The contribution of each source of variation was calculated using hierarchical variation partitioning (function varpart in the vegan package[59] in R). This procedure was repeated 100 times to obtain a median and confidence interval.

Assessing the relationship between biodiversity metrics and human impact

In further statistical analyses, we used the medians of biodiversity variables across sites per region. We related community completeness and other calculated biodiversity variables (alpha diversity, beta diversity, gamma diversity, dark diversity and species pool size) to the human footprint index from the year 2018[34]. The index ranges from 0 to 50 and is calculated from eight components (human population density, electric infrastructure, railways, roads, navigable waterways, the extent of built-up land, pastures and croplands). The resolution of the human influence data layers was 100 m, and we calculated average values over various spatial extents around the centre of each region (radii 10 km, 50 km, 100 km, 200 km, 300 km and 400 km). The averaging did not include areas representing water bodies. Because all regions included at least some areas less affected by humans, the total range of the averaged human footprint index values used in our analyses was somewhat lower than the maximum value. To test how well our sampled regions captured global variation in the human footprint index, we generated 500 random points worldwide using the discrete global grid system (which maintains uniform point density across the globe). From random points, we omitted glaciated regions of Antarctica and Greenland. We averaged the human footprint index in the surroundings of these random points in the same manner as we did with our empirical data. This revealed a high degree of correspondence between the average human footprint index ranges around sampled and randomly generated points at different scales: at radii of 50 km (sampled range 1.1–25.4, random 0.0–24.5), 200 km (sampled range 0.3–20.7, random 0.0–20.7) and 400 km (sampled 0.2–17.7, random 0.1–16.6).

To account for natural processes affecting community completeness, we included environmental variables in the multiple linear regression models. We used mean annual temperature and annual precipitation from the CHELSA database (resolution 1 km)[60,61], soil pH, organic carbon content, sand fraction proportion from SoilGrids (resolution 250 m)[62] and the topographic ruggedness of the terrain (resolution 250 m)[63]. Environmental factors were averaged within a 100-km radius to describe the broader region and consolidated through principal component analysis (PCA). For PCA, variables with only positive values were log-transformed if this resulted in a distribution closer to normal, and all variables were standardized. We kept the four first principal components, which described more than 90% of the variation. The first component was positively correlated with soil organic carbon content, acidity and precipitation; the second with temperature; the third with soil sand content; and the fourth with topographic ruggedness (Extended Data Fig. 5).

We fitted both linear and nonlinear (generalized additive models, function gam in the R package; ref. 64) models, incorporating the 116 regions in which all vascular plants were sampled. The estimates of the human footprint index at the different spatial scales were inherently strongly related to each other. Therefore, we constructed models for each scale at which the human footprint index (or its components) was averaged. We examined which scales produced the best models ($\Delta AIC < 2$) and selected the smallest scale, which is most directly related to the study region. We compared linear and nonlinear models using spatial block validation, implemented in the R package blockCV (ref. 65). We used fivefold cross-validation across hexagons (Extended Data Fig. 2). To estimate the variation in model predictive power we further implemented a bootstrap approach[66] by selecting bootstrap samples within each fold and then performing cross-validation. We used the normalized root mean square error (normalized by minimum and maximum values) to compare the predictive error of linear and nonlinear models, and found that linear models had much lower error in test sets (around 20% of the range compared with around 40% of the range; see Extended Data Fig. 8). Linear models were therefore used as a more general option.

We report the results of the best linear model (the smallest spatial scale at which $\Delta AIC < 2$) for each biodiversity metric and note significant relationships ($P < 0.05$). We used the variance inflation factor (VIF) to confirm that correlations between environmental gradients and human impact (Extended Data Fig. 5) were not confounding in the models (VIF < 2). We applied type III model testing. Consequently, the effect of human impact was tested only after the environmental effects were accounted for. We visualized the results of the fitted models in terms of how the predictor variable human footprint index affects the outcome of community completeness using the visreg function and package[67] in R. Model summary tables can be found in Extended Data Tables 1 and 2.

Besides the human footprint index from 2018, we also examined whether including change in the human footprint index during recent years improved the model[68]. Specifically, we tested whether a model including human footprint index change yielded a lower AIC value (by more than two units) compared with the model without change. We derived the measure of human footprint index change from a source that used a consistent methodology[69] during a temporal range 2000–2013. Change in human footprint index was quantified as log(human footprint index value from 2013/human footprint index value from 2000).

We tested whether community completeness is better described by certain quantiles of the human footprint index at different scales around study regions. Compared with the mean, considering quantiles allowed us to determine the extent to which it is important to maintain a certain proportion of area with lower human influence. We compared models incorporating as predictor variables the 10–90% quantiles of

the human footprint index using AIC and recorded cases in which the quantiles yielded a better model than the mean (models with AIC lower by more than two units were considered superior).

We also tested the robustness of the relationship between community completeness and the human footprint index by looking at statistical interactions between human influence and the environment, indirect effects, the role of sampling scale, alien or rare species; by including areas in which only woody species were recorded and considering forest cover in regions; and by examining the effect of geographically uneven sampling (see Supplementary Methods).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data supporting the findings of this study, along with the R scripts to handle them, can be found in Figshare: https://doi.org/10.6084/m9.figshare.25158059 (ref. 70). We also used published data for the human footprint index[34,69]; from the CHELSA database[60,61] for annual mean temperature and annual precipitation; from SoilGrids[62] for soil pH, organic carbon content and sand fraction proportion; and from the Geomorpho90m database[63] for the topographic ruggedness of the terrain.

55. Nathan, R. et al. Mechanisms of long-distance seed dispersal. *Trends Ecol. Evol.* **23**, 638–647 (2008).
56. Pollock, L. J. et al. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.* **5**, 397–406 (2014).
57. Chao, A. et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
58. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).
59. Oksanen, J. et al. vegan: Community Ecology Package. R version 4.2.2 https://CRAN.R-project.org/package=vegan (2022).
60. Karger, D. N. et al. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122 (2017).
61. Karger, D. N. et al. Data from: Climatologies at high resolution for the earth's land surface areas. *Dryad Digital Repository* https://doi.org/10.5061/dryad.kd1d4 (2017).
62. Hengl, T. et al. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* **12**, e0169748 (2017).
63. Amatulli, G., McInerney, D., Sethi, T., Strobl, P. & Domisch, S. Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. *Sci. Data* **7**, 162 (2020).
64. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. B* **73**, 3–36 (2011).
65. Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Arroita, G. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **10**, 225–232 (2019).
66. Fu, W. J., Carroll, R. J. & Wang, S. Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics* **21**, 1979–1986 (2005).
67. Breheny, P. & Burchett, W. Visualization of regression models using visreg. *R J.* **9**, 56–71 (2017).
68. Di Marco, M., Venter, O., Possingham, H. P. & Watson, J. E. M. Changes in human footprint drive changes in species extinction risk. *Nat. Commun.* **9**, 4621 (2018).
69. Williams, B. A. et al. Change in terrestrial human footprint drives continued loss of intact ecosystems. *One Earth* **3**, 371–382 (2020).
70. Pärtel, M. et al. Supporting data for 'Global impoverishment of natural vegetation revealed by dark diversity'. *Figshare* https://doi.org/10.6084/m9.figshare.25158059 (2025).

Acknowledgements We thank the Estonian Research Council (PRG609, PRG1065, PSG293 and PRG2142) and the Estonian Ministry of Education and Research (Centre of Excellence AgroCropFuture, TK200) for funding the DarkDivNet steering committee, and we thank the many students and volunteers who helped with the fieldwork. Acknowledgements related to specific authors can be found in the Supplementary Notes.

Author contributions The steering committee members (M.P., R. Tamme, C.P.C., K.R., M. Moora and M.Z.) initiated and coordinated DarkDivNet, analysed the data and wrote the manuscript. The advisory board members (J.A.B., A.C., M. Chytrý, F.B., O.E., S. Harrison, R.J.L., A.T.M., M.Ö. and J.N.P.) contributed to the study design and to the initial versions of the manuscript. V.A., D.A., Z.A., I.A., F.M.A., M.D.B., M.B., Z.B., N.B., K. Bergholz, K. Birkeli, I.B., J.M.B.-M., K.J.B., L.B.-M., B.B., P.H.S.B., F.Q.B., C.B., C.G.B., G.B., J.F.C., J.A.C., G.C., M. Carbognani, C.C., B.E.L.C., R.C., J.S.C., J.M.C., S.A.O.C., J.Č., M. Dairel, M.D.F., A.D., J. Davison, B.D., S.D.V., I.D., J. Dengler, J. Dolezal, X.D., M. Dvorsky, H.E., L.E., D.E., A.E., F.E., G.F., E.F., F.F., E.F.-P., A. Ferrara, A. Fidelis, M. Fischer, M. Flagmeier, T.G.W.F., L.H.F., J.F., F.F.F., B. Garris, H.W.G., M.A.G., G.G.G., A.G.-R., M.K.G., M.G., R.G., I.G., J.G., B. Güler, Y.G., S Haesen, M.H., R.H.H., T.T.H., R.H., Y.H., J.T.H., D.I., R.I., N.I., S.D.H.I., E.J., F. Jansen, F. Jeltsch, A.J., B.J.-A., M.J., M.H.J., S.K., N. Katal, A.K., B.I.K., A.A.K., K.J.K., M.K., E.K., L.K., N. Koroleva, K.A.K., M.V.K., Ł.K., L. Laanisto, H. Lager, V.L., R.G.L., J.J.L., L. Li, A.L., H. Liu, K. Liu, X.L., M.E.L.-B., K. Ludewig, K. Lukács, J.L.-M., P.M., M. Marignani, R.M., T.M., J.E.M., K.M., D.M., E.M.-J., I.M., L.M., M.M.-R., R.A.M., S.M.N., L.N., S.N., A.N., P.N., T.O., V.G.O., K.L.P., B. Paganeli, B. Peco, A.M.L.P., A.P.-H., P.L.P., A.P., G.P., P.A.P.-Á., J.P., H.C.P., V.E.P., D.R., S.R., T.R., M.R., A.A.R., A.B.R., D.A.R.G., R.R., R.E.R., L.R., J. Sádlo, K. Salimbayeva, R.S.D., K. Sanchir, C.S., J.D.S., U.S., J. Schrader, N.L.S., G. Sellan, J.M.S.-D., G. Silan, H.S., N.S., J. Sonkoly, K. Štajerova, I.S., S.Ś., A.J.T., F.M.T., R. Tarifa, P. Tejero, D.K.T., M. Tholin, R.S.T., Y.T., A. Tokaryuk, C.T., M. Tomaselli, E.T., P. Török, B. Tóthmérész, A. Toussaint, B. Touzard, D.P.F.T., J.L.T., S.T., E.V., M. Valerio, O.V., K.V.M., V.V., J.V., R.V., M. Vítkova, M. Vojík, A.H., J.O., V.W., J.-Z.W., C.-J.W., S.A.W., L.W., T.W., S.X. and O.Z.M. provided data and were involved in the interpretation of the results and manuscript preparation.

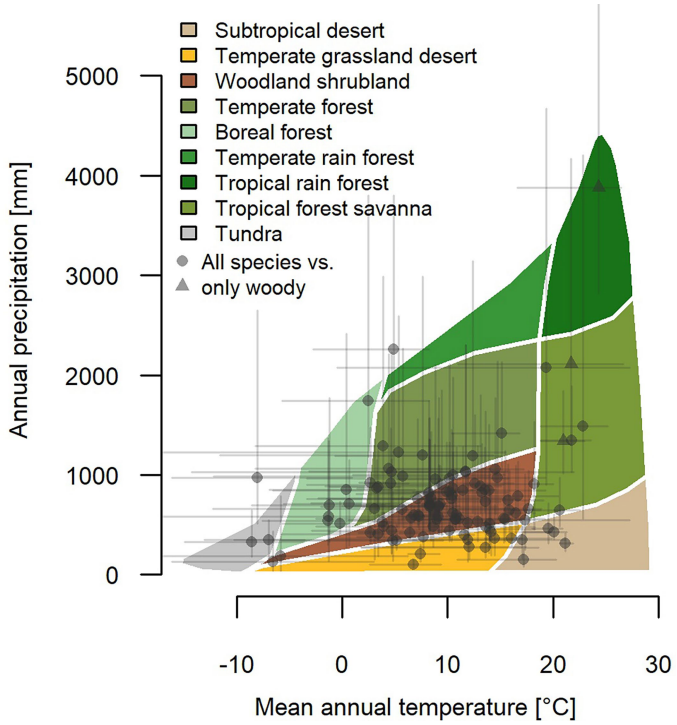
Competing interests The authors declare no competing interests.

Additional information

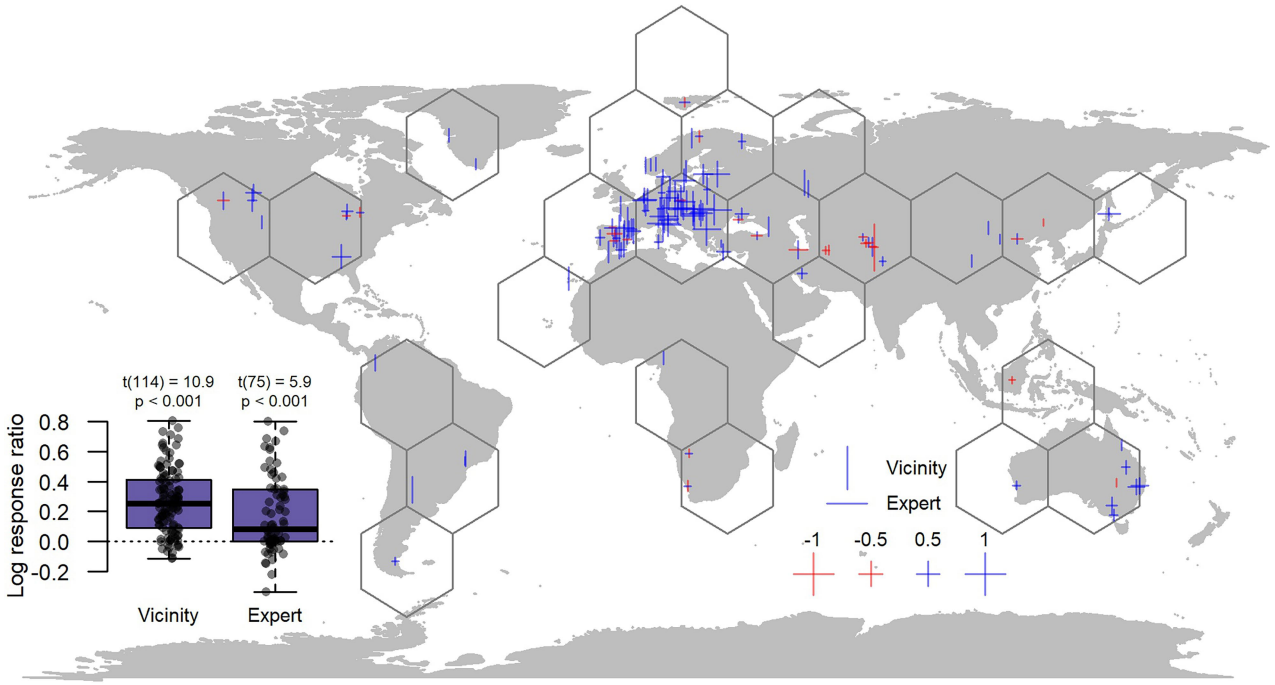
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-08814-5.

Correspondence and requests for materials should be addressed to Meelis Pärtel.

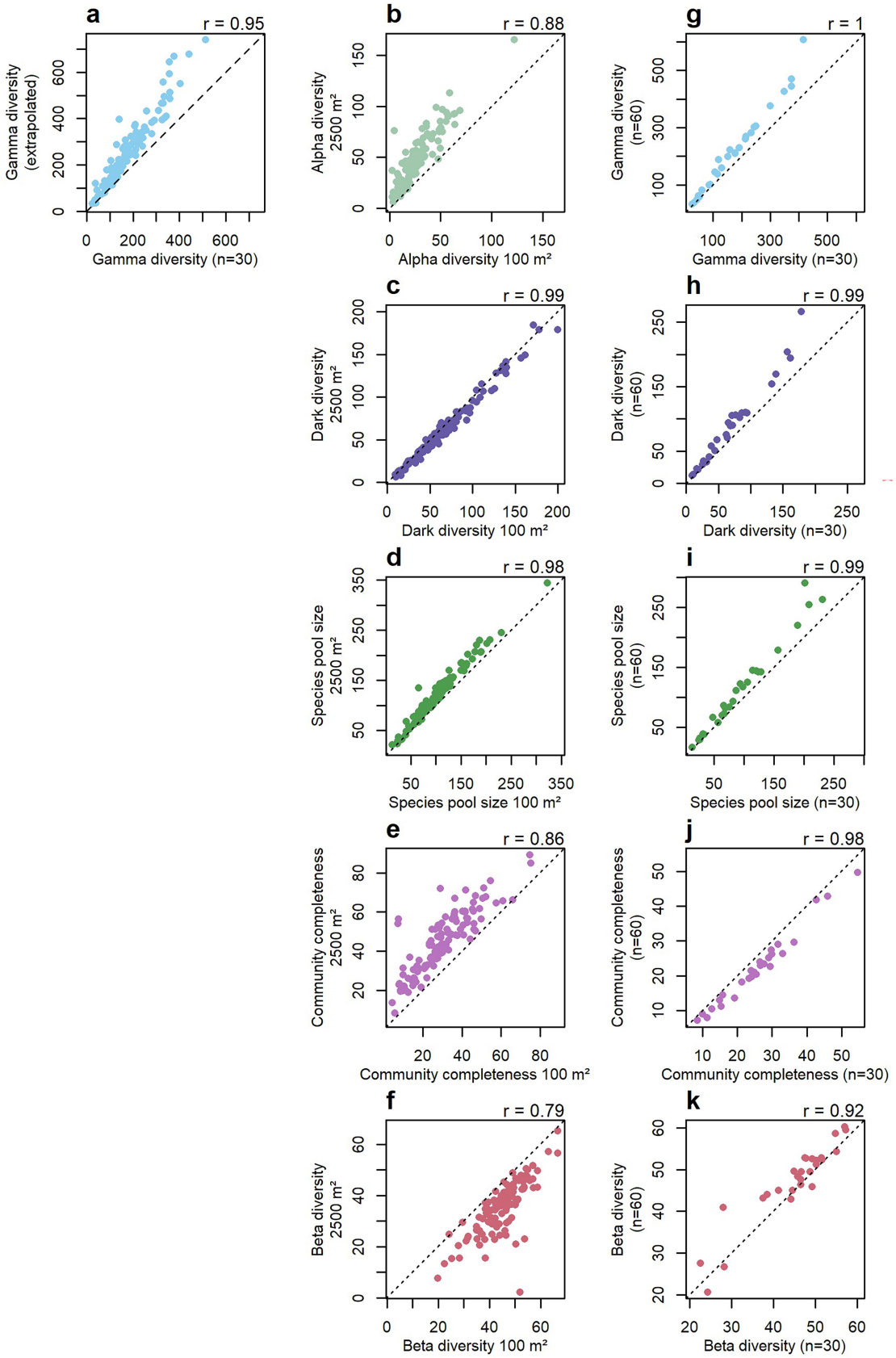
Peer review information *Nature* thanks Marta Jarzyna, Alejandro Ordonez and Moreno Di Marco for their contribution to the peer review of this work.



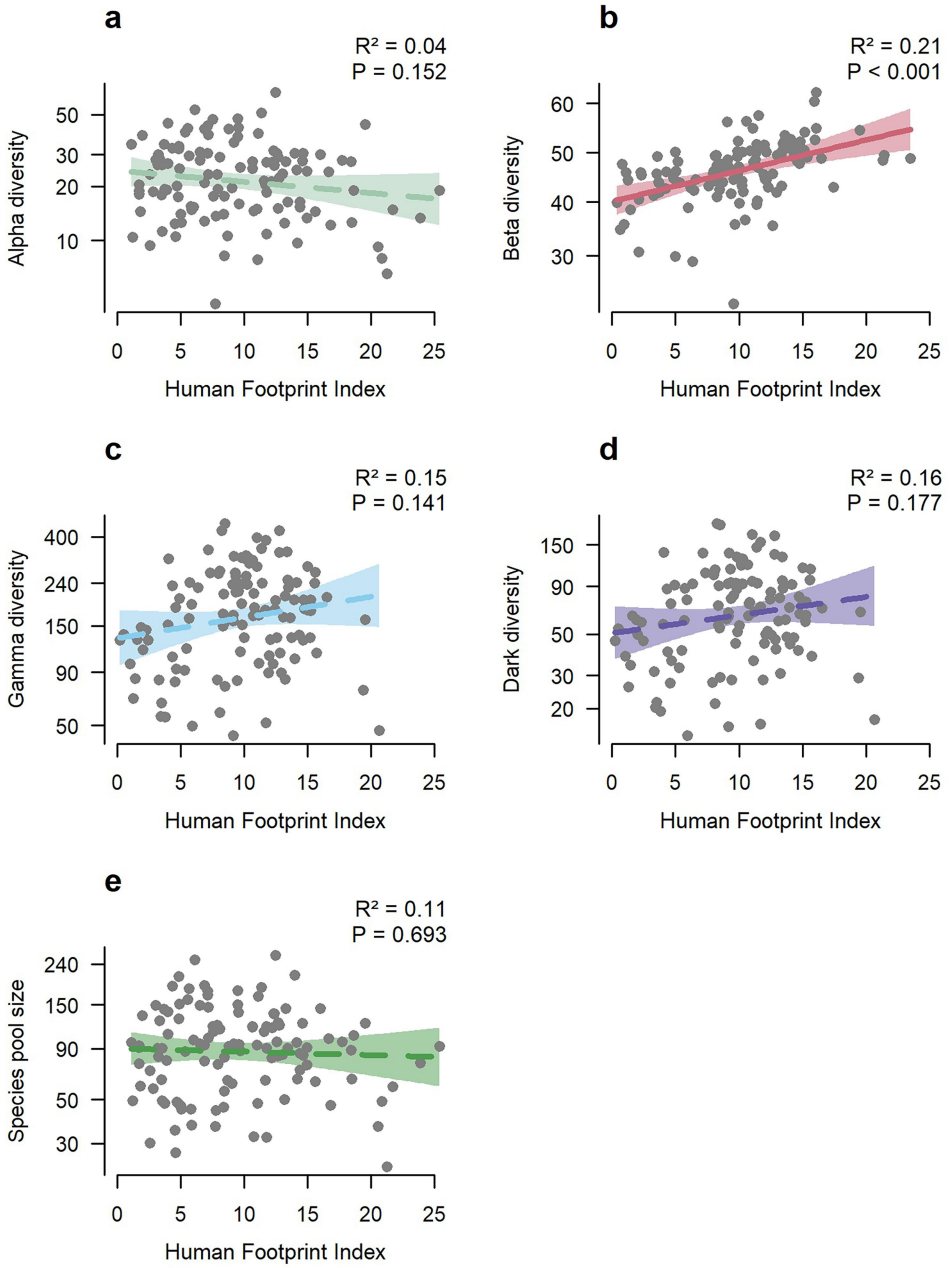
Extended Data Fig. 1 | Distribution of the 119 DarkDivNet study regions in relation to mean annual temperature and annual precipitation. Lines indicate ranges within a radius of 100 km. Approximate broad biomes are shown. Triangles indicate regions in which only woody species were sampled.



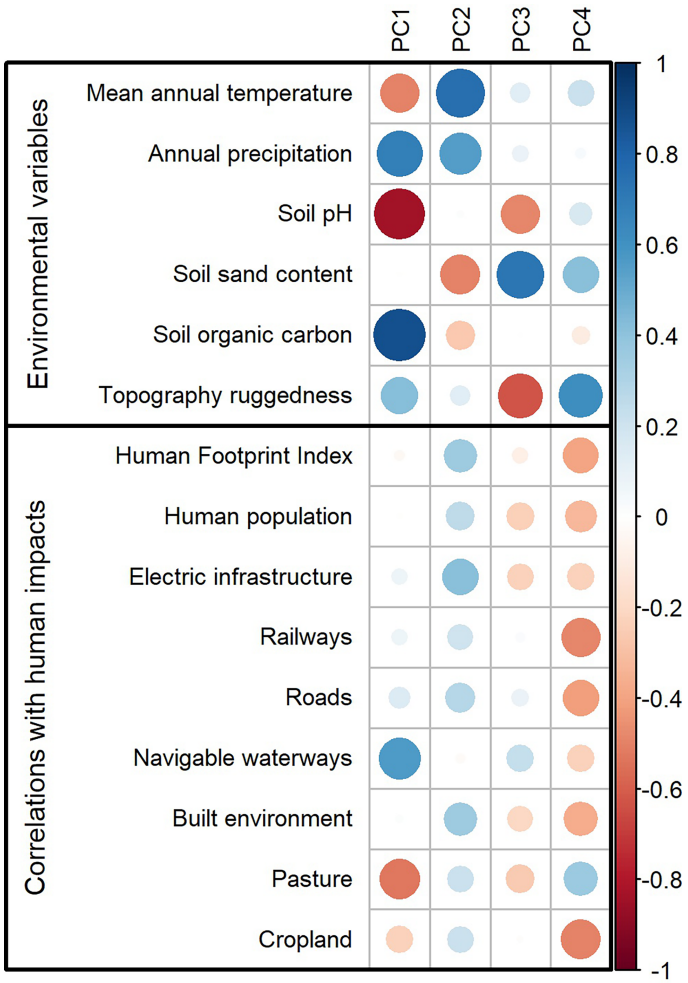
Extended Data Fig. 2 | Using independent data to test the dark diversity method that relies on species co-occurrences to estimate the ecological suitability of absent species. We used two tests. In the Vicinity test, we examined whether species absent from the site (100 m^2) but present in the immediate vicinity (2500 m^2) have higher estimated suitabilities than absent species found further away. The sample area and vicinity area are assumed to share relatively similar ecological conditions. In the Expert test, we compared whether species absent from the site but assessed by expert opinion to be ecologically suitable (i.e., belong to the site-specific species pool) have higher calculated suitabilities than those absent species that were evaluated as unsuitable. In both cases, we calculated the log response ratio of the mean suitability of species in the respective groups. Positive log response ratios indicate agreement between assessments of suitability calculated from co-occurrences and from the independent information considered in the tests. The length of the lines (vertical for the Vicinity test and horizontal for the Expert test) shown at study region locations indicates the magnitude of the log response ratio; negative values are in red and positive values are in blue. Both tests comprised data from a subset of study regions. The box plot on the left (centre line, median; box limits, upper and lower quartiles; whiskers, the range, excluding outlying points that exceed the quartiles by more than 1.5× the interquartile range) shows the results of single-sample two-sided t-tests (difference from zero), with log response ratios significantly larger than zero in both cases, $n = 115$ regions for the Vicinity test and $n = 76$ regions for the Expert test. Hexagons on the map (made with Natural Earth; free vector and raster map data; https://www.naturalearthdata.com/) delimit the spatial blocks used in cross-validation.



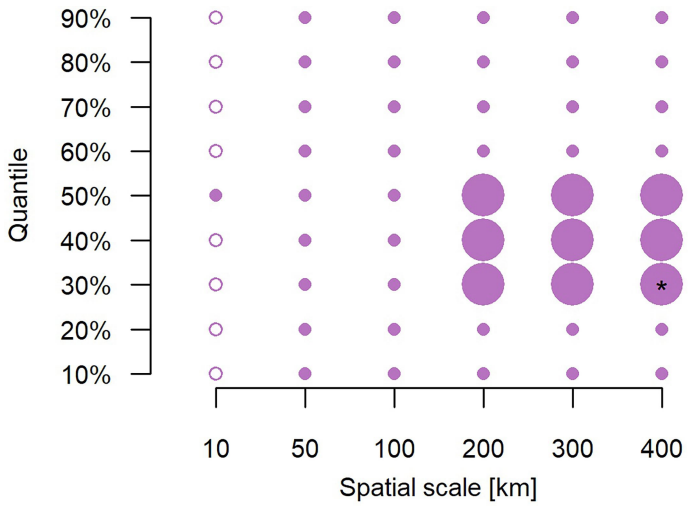
Extended Data Fig. 3 | Bivariate plots of biodiversity metrics derived using alternative sampling methods. a, Gamma diversity from 30 sites compared with extrapolation up to complete sampling coverage. **b**–**f**, Sites described in a 2500 m^2 area in addition to the DarkDivNet standard of 100 m^2. **g**–**k**, Biodiversity metrics when 60 sites were used to estimate co-occurrences in addition to the DarkDivNet standard of 30. The scatter plots show mean values for regions where the respective sampling scheme was applied ($n = 119$ regions for **a**, $n = 116$ regions for **b**–**f** and $n = 27$ for **g**–**k**). The 1:1 lines are shown as diagonals. Estimates of Spearman correlation for each comparison are shown above the panels. Comparisons where the alternative sampling method did not influence the metric (i.e. gamma diversity when using a larger sample area or alpha diversity when using more sites) are not shown.



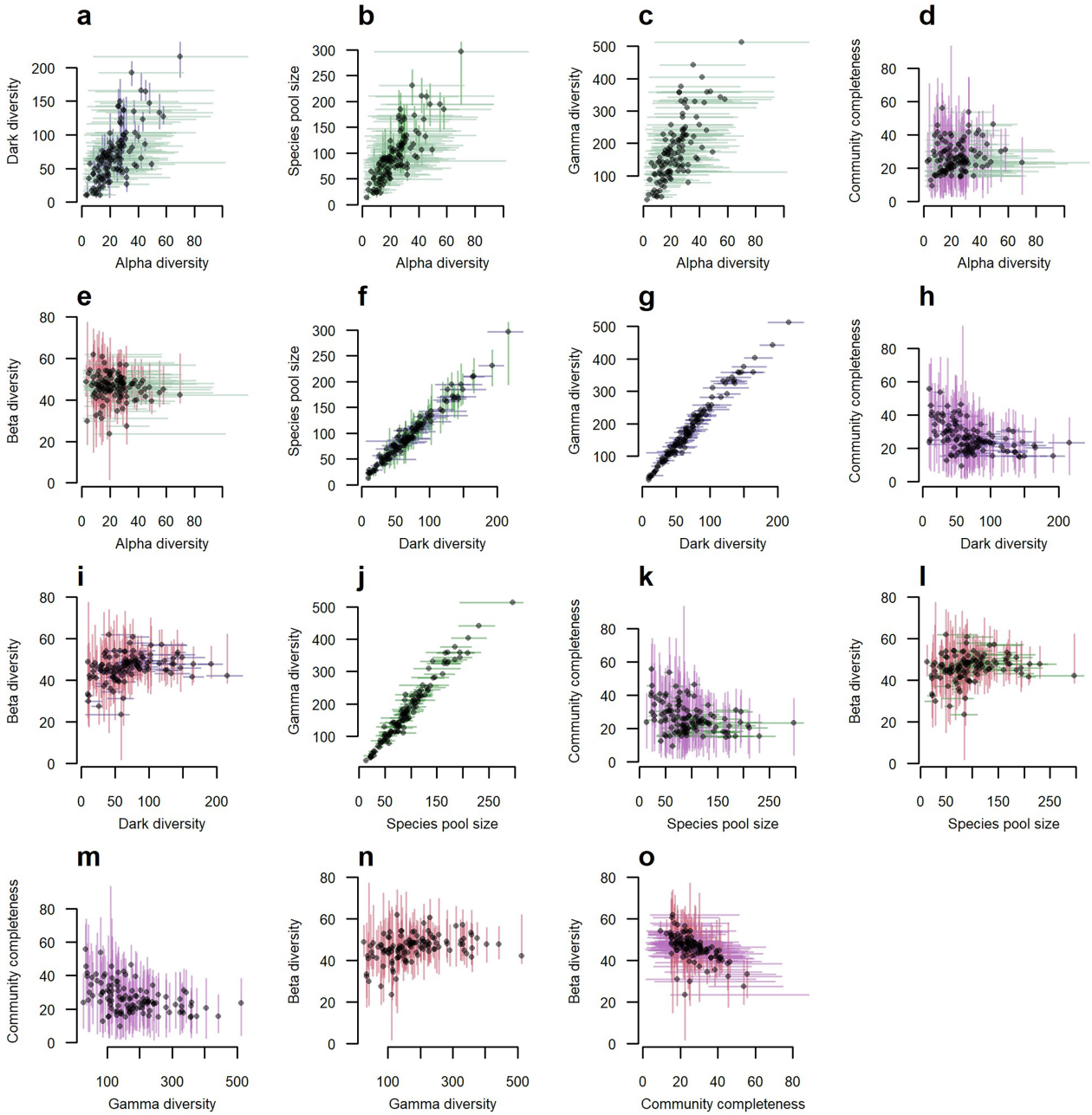
Extended Data Fig. 4 | Relationships between biodiversity metrics and the human footprint index in the surrounding regions. a, Alpha diversity. **b**, Beta diversity. **c**, Gamma diversity. **d**, Dark diversity. **e**, Species pool size. A similar graph for community completeness is shown in Fig. 2a. For each metric, the relationships from the spatial scale producing the best multiple linear regression model is shown (n = 116 regions, see Extended Data Table 1 for details of the models). The prediction lines are shown with 95% confidence intervals. The solid line indicates a significant relationship (two-tailed $p < 0.05$); the dashed lines indicate non-significant trends. Note that the range of the human footprint index varies when averaged at different spatial scales. Diversity values on y-axes are back-transformed from log or logit scales.



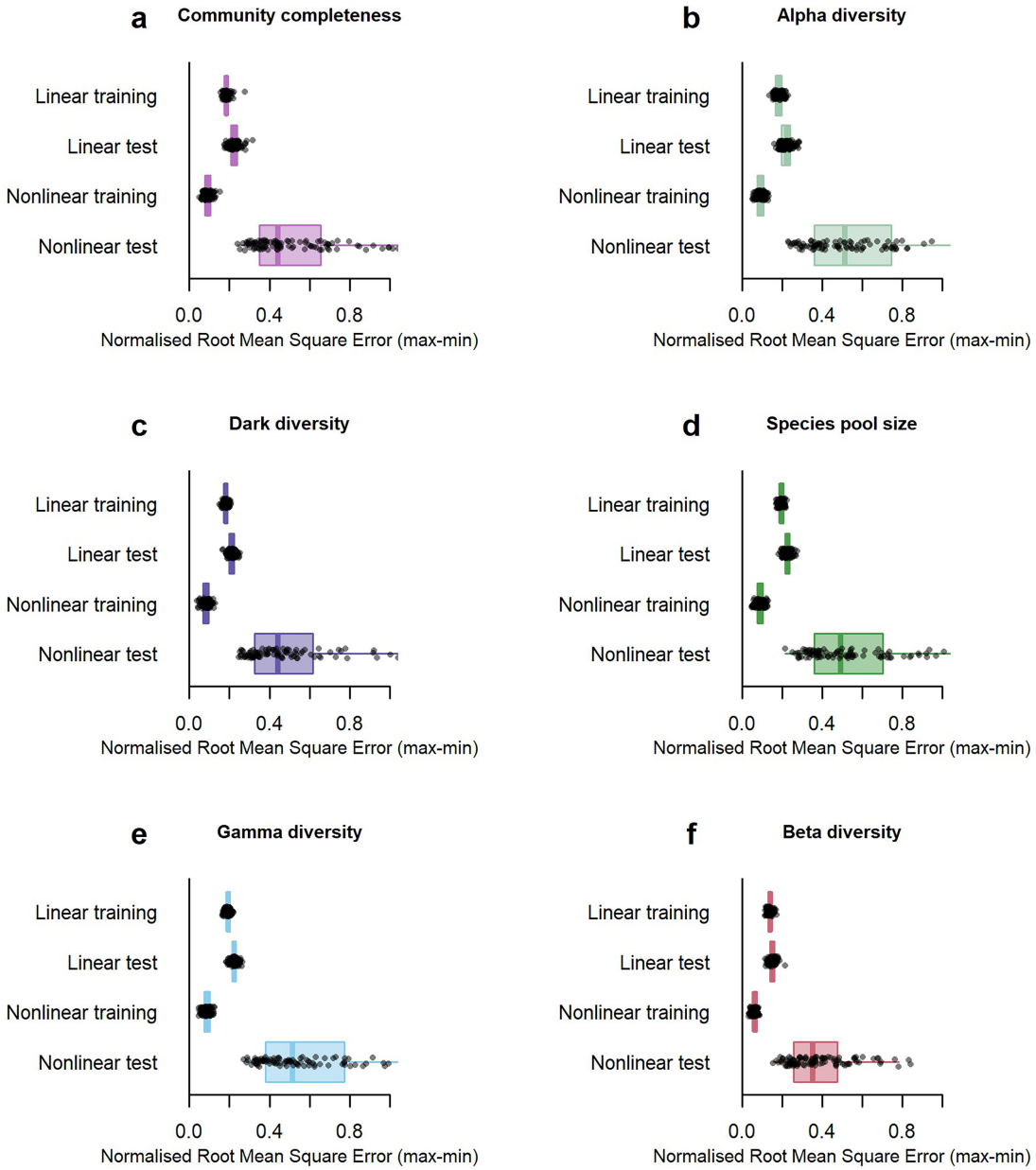
Extended Data Fig. 5 | Correlations between four principal components and explanatory variables. Correlations with raw environmental variables (used in the PCA; top) and the human footprint index or its components (not used in the PCA; bottom).



Extended Data Fig. 6 | Improvement in multiple linear regression models describing community completeness provided by including quantiles of the human footprint index values found in regions at various spatial scales. Filled symbols indicate significant relationships (n = 116 regions, two-tailed $p < 0.05$), and the large symbols indicate models where the Akaike information criterion (AIC) is lower than the minimal value among the models using the mean human footprint index value (Extended Data Table 1). The asterisk indicates a combination of quantile and spatial scale that yielded a considerably better model (AIC value lower by more than 2 units).



Extended Data Fig. 7 | Bivariate relationships between diversity metrics. a-**e**, Alpha diversity. **a**, **f**-**i**, Dark diversity. **b**, **f**, **j**-**l**, Species pool size. **c**, **g**, **j**, **m**-**o**, Gamma diversity. **d**, **h**, **k**, **m**, **o**, Community completeness. **e**, **i**, **l**, **n**, **o**, Beta diversity. Lines indicate variation within regions (99% quantiles, i.e. omitting outliers), crossing at median values within regions. Several metrics are inherently related (see Fig. 1), and strong relationships are expected. However, variation within regions can be large, indicating the importance of site-specific metrics. The colours of lines reflect the different biodiversity metrics (see Fig. 1) to facilitate comparisons across panels.



Extended Data Fig. 8 | Normalized root mean square error values from linear and nonlinear models predicting various biodiversity metrics. **a**, Community completeness. **b**, Alpha diversity. **c**, Dark diversity. **d**, Species pool size. **e**, Gamma diversity. **f**, Beta diversity. We used fivefold spatial cross-validation (see Extended Data Fig. 2) with bootstrapping to estimate the variation. The box plots (centre line = median; box limits = upper and lower quartiles; whiskers = the range, excluding outlying points that exceed the quartiles by more than 1.5× the interquartile range) show that while the nonlinear models had lower errors for the training set, the test data were predicted with lower error by the linear models.

Extended Data Table 1 | Statistical summary tables from multiple linear regression models in which biodiversity metrics were modelled in relation to the human footprint index while also considering environmental variables (principal components)

Study variable	Descriptor	Sum Sq.	df	F	P-value	St. coef.
Community completeness [a]	**PC1**	**1.98**	**1**	**13.82**	**<0.001**	**-0.30**
	PC2	0.06	1	0.44	0.507	-0.06
The best model: 300 km	PC3	0.30	1	2.06	0.154	-0.12
Max VIF value for the best model: 1.66	PC4	0.35	1	2.41	0.123	-0.14
R^2 of the best model: 0.28	**Human Footprint Index**	**2.37**	**1**	**16.55**	**<0.001**	**-0.42**
	Residuals	15.78	110			
Alpha diversity [b]	PC1	0.14	1	0.57	0.453	0.07
	PC2	0.64	1	2.52	0.115	0.16
The best model: 10 km	PC3	0.05	1	0.20	0.656	-0.04
Max VIF value for the best model: 1.33	PC4	0.18	1	0.71	0.402	-0.09
R^2 of the best model: 0.04	Human Footprint Index	0.52	1	2.08	0.152	-0.16
	Residuals	27.78	110			
Dark diversity [c]	**PC1**	**3.47**	**1**	**10.26**	**0.002**	**0.28**
	PC2	1.17	1	3.46	0.065	0.20
The best model: 200 km	PC3	0.04	1	0.12	0.733	0.03
Max VIF value for the best model: 1.75	PC4	0.01	1	0.04	0.845	0.02
R^2 of the best model: 0.16	Human Footprint Index	0.63	1	1.85	0.177	0.16
	Residuals	37.24	110			
Species pool size [d]	**PC1**	**1.84**	**1**	**6.88**	**0.010**	**0.24**
	PC2	**1.91**	**1**	**7.15**	**0.009**	**0.26**
The best model: 10 km	PC3	0.01	1	0.04	0.835	0.02
Max VIF value for the best model: 1.33	PC4	0.09	1	0.34	0.562	-0.06
R^2 of the best model: 0.11	Human Footprint Index	0.04	1	0.16	0.693	-0.04
	Residuals	29.38	110			
Gamma diversity [e]	**PC1**	**2.39**	**1**	**8.05**	**0.005**	**0.25**
	PC2	0.90	1	3.05	0.084	0.18
The best model: 200 km	PC3	0.00	1	0.00	0.974	0.00
Max VIF value for the best model: 1.75	PC4	0.00	1	0.01	0.907	0.01
R^2 of the best model: 0.15	Human Footprint Index	0.65	1	2.20	0.141	0.17
	Residuals	32.59	110			
Beta diversity [f]	PC1	0.20	1	3.78	0.054	0.17
	PC2	0.00	1	0.00	0.963	0.00
The best model: 100 km	PC3	0.04	1	0.71	0.400	-0.07
Max VIF value for the best model: 1.78	PC4	0.06	1	1.05	0.308	0.10
R^2 of the best model: 0.21	**Human Footprint Index**	**0.93**	**1**	**17.34**	**<0.001**	**0.47**
	Residuals		110			

[a] Human impact had significant effects ($p < 0.05$) at scales 50, 100, 200, 300, 400 km; ΔAIC values across scales: 10 km: 13.9, 50 km: 10.8, 100 km: 6.9, 200 km: 2.8, 300 km: 1.2, 400 km: 0. [b] Human impact not significantly related at any studied scale ($p > 0.05$); ΔAIC values across scales: 10 km: 1.9, 50 km: 0, 100 km: 0.7, 200 km: 2.1, 300 km: 2.7, 400 km: 2. [c] Human impact not significantly related at any studied scale ($p > 0.05$); ΔAIC values across scales: 10 km: 2.9, 50 km: 2.9, 100 km: 2.4, 200 km: 0.9, 300 km: 0, 400 km: 0.4. [d] Human impact not significantly related at any studied scale ($p > 0.05$); ΔAIC values across scales: 10 km: 0.8, 50 km: 0.7, 100 km: 1, 200 km: 0.6, 300 km: 0, 400 km: 0.3. [e] Human impact not significantly related at any studied scale ($p > 0.05$); ΔAIC values across scales: 10 km: 3.2, 50 km: 3.2, 100 km: 2.6, 200 km: 0.9, 300 km: 0, 400 km: 0.6. [f] Human impact had significant effects ($p < 0.05$) at scales 50, 100, 200, 300, 400 km; ΔAIC values across scales: 10 km: 14.5, 50 km: 7.4, 100 km: 1.2, 200 km: 0, 300 km: 1.7, 400 km: 2.6.

Type III tests were used to estimate the two-tailed significance of each independent variable, n=116 regions. For each model output, information is provided about the spatial range of human influence that produced the best model according to the Akaike information criterion (the smallest scale for which ΔAIC<2), variance inflation factors and model R^2 values (grey boxes).

Extended Data Table 2 | Statistical summary tables from multiple linear regression models in which community completeness was modelled in relation to the human footprint index and its components while also considering environmental variables (principal components)

Study variable	Descriptor	Sum Sq.	df	F	P-value	St. coef.
Community completeness [a]	**PC1**	**1.98**	**1**	**13.82**	**<0.001**	**-0.30**
	PC2	0.06	1	0.44	0.507	-0.06
The best model: 300 km	PC3	0.30	1	2.06	0.154	-0.12
Max VIF value for the best model: 1.66	PC4	0.35	1	2.41	0.123	-0.14
R^2 of the best model: 0.28	**Human Footprint Index**	**2.37**	**1**	**16.55**	**<0.001**	**-0.42**
	Residuals	15.78	110			
Community completeness [b]	**PC1**	**1.90**	**1**	**12.12**	**0.001**	**-0.30**
	PC2	**0.63**	**1**	**4.05**	**0.047**	**-0.18**
The best model: 300 km	PC3	0.62	1	3.93	0.050	-0.18
Max VIF value for the best model: 1.42	PC4	0.04	1	0.25	0.618	-0.04
R^2 of the best model: 0.21	**Human population**	**0.92**	**1**	**5.88**	**0.017**	**-0.25**
	Residuals	17.24	110			
Community completeness [c]	**PC1**	**1.30**	**1**	**8.65**	**0.004**	**-0.25**
	PC2	0.21	1	1.40	0.239	-0.11
The best model: 300 km	PC3	0.57	1	3.81	0.053	-0.17
Max VIF value for the best model: 1.41	PC4	0.02	1	0.14	0.714	-0.03
R^2 of the best model: 0.25	**Electric infrastructure**	**1.80**	**1**	**12.03**	**0.001**	**-0.34**
	Residuals	16.34	109			
Community completeness [d]	**PC1**	**1.36**	**1**	**9.52**	**0.003**	**-0.26**
	PC2	0.52	1	3.66	0.058	-0.16
The best model: 200 km	PC3	0.12	1	0.86	0.356	-0.08
Max VIF value for the best model: 1.4	PC4	0.35	1	2.41	0.123	-0.14
R^2 of the best model: 0.28	**Railways**	**2.39**	**1**	**16.68**	**<0.001**	**-0.39**
	Residuals	15.77	110			
Community completeness [e]	**PC1**	**0.80**	**1**	**5.73**	**0.018**	**-0.20**
	PC2	0.07	1	0.48	0.490	-0.06
The best model: 400 km	PC3	0.08	1	0.54	0.462	-0.06
Max VIF value for the best model: 1.65	PC4	0.38	1	2.71	0.103	-0.15
R^2 of the best model: 0.30	**Roads**	**2.81**	**1**	**20.15**	**<0.001**	**-0.46**
	Residuals	15.35	110			
Community completeness [f]	**PC1**	**2.45**	**1**	**15.15**	**<0.001**	**-0.34**
	PC2	**1.65**	**1**	**10.16**	**0.002**	**-0.28**
The best model: 10 km	PC3	0.32	1	2.00	0.160	-0.12
Max VIF value for the best model: 1.07	PC4	0.01	1	0.07	0.788	0.02
R^2 of the best model: 0.18	Navigable waterways	0.34	1	2.09	0.151	0.13
	Residuals	17.82	110			
Community completeness [g]	**PC1**	**1.62**	**1**	**10.79**	**0.001**	**-0.28**
	PC2	0.46	1	3.08	0.082	-0.16
The best model: 300 km	PC3	0.42	1	2.77	0.099	-0.14
Max VIF value for the best model: 1.38	PC4	0.15	1	1.03	0.313	-0.09
R^2 of the best model: 0.24	**Built environment**	**1.69**	1	**11.30**	**0.001**	**-0.33**
	Residuals	16.47	110			
Community completeness [h]	**PC1**	**1.11**	**1**	**6.87**	**0.010**	**-0.25**
	PC2	**1.48**	**1**	**9.20**	**0.003**	**-0.26**
The best model: 10 km	PC3	0.17	1	1.07	0.303	-0.09
Max VIF value for the best model: 1.34	PC4	0.00	1	0.02	0.880	-0.01
R^2 of the best model: 0.19	Pasture	0.41	1	2.56	0.113	0.16
	Residuals	17.74	110			
Community completeness [i]	**PC1**	**3.32**	**1**	**21.95**	**<0.001**	**-0.42**
	PC2	**0.75**	**1**	**4.96**	**0.028**	**-0.19**
The best model: 100 km	PC3	0.12	1	0.82	0.368	-0.08
Max VIF value for the best model: 2.02	PC4	0.51	1	3.34	0.070	-0.20
R^2 of the best model: 0.24	**Cropland**	**1.52**	**1**	**10.08**	**0.002**	**-0.38**
	Residuals	16.63	110			

[a] Human impact had significant effects ($p < 0.05$) at scales 50, 100, 200, 300, 400 km; ΔAIC values across scales: 10 km: 13.9, 50 km: 10.8, 100 km: 6.9, 200 km: 2.8, 300 km: 1.2, 400 km: 0. [b] Human impact had significant effects ($p < 0.05$) at scales 200, 300, 400 km; ΔAIC values across scales: 10 km: 6.6, 50 km: 5.1, 100 km: 4.5, 200 km: 2.6, 300 km: 1.5, 400 km: 0. [c] Human impact had significant effects ($p < 0.05$) at scales 200, 300, 400 km; ΔAIC values across scales: 10 km: 11, 50 km: 10.2, 100 km: 9.3, 200 km: 5.6, 300 km: 0.4, 400 km: 0. [d] Human impact had significant ($p < 0.05$) effects at all studied scales; ΔAIC values across scales: 10 km: 10.9, 50 km: 5.4, 100 km: 2.4, 200 km: 0.4, 300 km:0.2, 400 km: 0. [e] Human impact had significant ($p < 0.05$) effects at all studied scales; ΔAIC values across scales: 10 km: 14.5, 50 km: 11, 100 km: 7.5, 200 km: 5, 300 km: 3, 400 km: 0. [f] Human impact not significantly related at any studied scale ($p > 0.05$); ΔAIC values across scales: 10 km: 0.2, 50 km: 0, 100 km: 1.9, 200 km: 1.3, 300 km: 0.8, 400 km: 1.6. [g] Human impact had significant effects ($p < 0.05$) at scales 100, 200, 300, 400 km. ΔAIC values across scales: 10 km: 10.1, 50 km: 9.1, 100 km: 8.1, 200 km: 5, 300 km: 1, 400 km: 0. [h] Human impact not significantly related at any studied scale ($p > 0.05$); ΔAIC values across scales: 10 km: 1.1, 50 km: 0, 100 km: 0.7, 200 km: 0.8, 300 km: 2.2, 400 km: 2. [i] Human impact had significant effects ($p < 0.05$) at scales 50, 100, 200, 300, 400 km; ΔAIC values across scales: 10 km: 10.2, 50 km: 6.3, 100 km: 1.6, 200 km: 0, 300 km: 4.2, 400 km: 5.1.

Type III tests were used to estimate the two-tailed significance of each independent variable, n=116 regions. For each model output, information is provided about the spatial range of human influence that produced the best model according to the Akaike information criterion (the smallest scale for which ΔAIC<2), variance inflation factors and model R^2 values (grey boxes).

nature portfolio

Corresponding author(s): Meelis Pärtel

Last updated by author(s): Jan 22, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	R version 4.2.2, packages Hmsc (3.0-13), iNEXT (3.0.0), blockCV (3.1-3), lavaan (0.6-16), mgcv (1.8-41), vegan (2.6-4), car (3.1-1), ape (5.7-1), sf (1.0-9), DarkDiv (0.3.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data supporting the findings of this study along with the R scripts to handle them can be found in Figshare: https://doi.org/10.6084/m9.figshare.25158059. We additionally used published data for the Human Footprint Index (refs 34,69), from the Chelsa database (refs 60,61) for annual mean temperature and annual

precipitation, from SoilGrids (ref. 62) for soil pH, organic carbon content, sand fraction proportion, and from the Geomorpho90m database (ref. 63) for the topographic ruggedness of the terrain.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Research is not involving human participants, their data, or biological material.
Reporting on race, ethnicity, or other socially relevant groupings	Research is not involving human participants, their data, or biological material.
Population characteristics	Research is not involving human participants, their data, or biological material.
Recruitment	Research is not involving human participants, their data, or biological material.
Ethics oversight	Research is not involving human participants, their data, or biological material.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The study is based on observational data from 5415 sample sites in 119 regions worldwide where natural and semi-natural vegetation has been described. Each region had at last 30 sample sites. Species co-occurrences within a region was used to estimate dark diversity -- species ecologically suitable for a study site which are absent. In the analyses, median biodiversity values of each region were used as replicates. Different biodiversity metrics were related to Human Footprint Index after accounting for covariates (climate, soil, topography). Several additional tests were done to evaluate the robustness of the findings.
Research sample	A study site was described by 100 m² (10 m × 10 m) vegetation plot where all vascular plant species were recorded. A subset of sites were also sampled by a 2500 m² (50 m × 50 m) sample plot to test methods and the sample scale effect. Explanatory variables were obtained from global databases: Human Footprint Index and its eight components (human population density, electric infrastructure, railways, roads, navigable waterways, the extent of built-up land, pastures, and croplands), climate data, soil parameters, and topography.
Sampling strategy	Sampling protocol was elaborated by the advisory board of the collaborative research network. Sample size was determined as the optimal balance between the amount of biodiversity information captured and feasibility in the field works. The decision to have a minimum of 30 samples per study region is based on simulation data (Carmona & Pärtel 2021 Global Ecology Biogeography).
Data collection	The data collection is based on a collaborative research network DarkDivNet where teams were invited to sample their own study regions. All data have been collected by authors of this work.
Timing and spatial scale	Sampling occurred during 2018-2023. Each sample site was defined as a 100 m² area in a natural or semi-natural vegetation within a region. A region represents the surrounding area of ca 300 km², defined by a circle of 20 km diameter with the available area influenced by geographical and practical limitations (coastline, private ownership and other restricted access).
Data exclusions	From primary analyzes we excluded three tropical regions where only woody species were sampled. The DarkDivNet protocol allowed only woody species sampling in hyperdiverse regions and theoretically some parameters (community completeness, beta diversity) is not much affected by inclusion of woody species only) but for robustness we only included regions where all vascular plant species were recorded. The excluded regions were still used in a separate analyses showing that results were same if we included them.
Reproducibility	Before the field work we agreed in a common sampling protocol which was available to all co-authors. If there were questions about some methodological details, then DarkDivNet steering committee advised how to keep consistent sampling across each study region.
Randomization	Randomization of sample regions were not feasible but potential covariates (climate, soil, topography) were included to statistical tests. The effect of human influence on biodiversity metrics was tested after accounting for the variation described by abiotic conditions (type III ANOVA). We further examined if the potential effect of global distribution of study regions using three methods:

(1) we looked the relationship between study variables in subsets where a single study region was iteratively selected from each represented ecoregion; (2) including factor "Europe" in the model since it was the most represented continent; (3) examining the spatial autocorrelation of model residuals (Moran I). In statistical tests randomization was used to have comparable data coverage across regions -- the same number of study sites were selected from each region iteratively and median of the outcome was used in the further analyses.

Blinding: Blinding was not applied since we had no treatments but observational data.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions: Field work was performed during the typical vegetation description time when most plant species were identifiable.

Location: Globally 119 regions.

Access & import/export: No collected samples were used in this study, only description done during the field work. Co-authors worked in their established study areas.

Disturbance: The field work included vegetation descriptions causing no long-lasting disturbance to the nature.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks: No seed stocks were used.

Novel plant genotypes: Novel plant genotypes were not produced.

Authentication: No seed stocks nor plant genotypes were used.