

Global hotspots for soil nature conservation

<https://doi.org/10.1038/s41586-022-05292-x>

Received: 17 November 2021

Accepted: 30 August 2022

Published online: 12 October 2022

 Check for updates

Carlos A. Guerra^{1,2,3}✉, Miguel Berdugo⁴, David J. Eldridge⁵, Nico Eisenhauer^{1,3}, Brajesh K. Singh^{6,7}, Haiying Cui^{8,9}, Sebastian Abades¹⁰, Fernando D. Alfaro^{10,11}, Adebola R. Bamigboye¹², Felipe Bastida¹³, José L. Blanco-Pastor¹⁴, Asunción de los Ríos¹⁵, Jorge Durán^{16,17}, Tine Grebenc¹⁸, Javier G. Illán¹⁹, Yu-Rong Liu²⁰, Thulani P. Makhalanyane²¹, Steven Mamet²², Marco A. Molina-Montenegro^{23,24}, José L. Moreno¹³, Arpan Mukherjee²⁵, Tina U. Nahberger¹⁸, Gabriel F. Peñaloza-Bojacá²⁶, César Plaza²⁷, Sergio Pico²⁸, Jay Prakash Verma²⁵, Ana Rey¹⁵, Alexandra Rodríguez¹⁶, Leho Tedersoo^{29,30}, Alberto L. Teixido³¹, Cristian Torres-Díaz³², Pankaj Trivedi³³, Juntao Wang⁶, Ling Wang⁸, Jianyong Wang⁸, Eli Zaady³⁴, Xiaobing Zhou³⁵, Xin-Quan Zhou²⁰ & Manuel Delgado-Baquerizo^{36,37}✉

Soils are the foundation of all terrestrial ecosystems¹. However, unlike for plants and animals, a global assessment of hotspots for soil nature conservation is still lacking². This hampers our ability to establish nature conservation priorities for the multiple dimensions that support the soil system: from soil biodiversity to ecosystem services. Here, to identify global hotspots for soil nature conservation, we performed a global field survey that includes observations of biodiversity (archaea, bacteria, fungi, protists and invertebrates) and functions (critical for six ecosystem services) in 615 composite samples of topsoil from a standardized survey in all continents. We found that each of the different ecological dimensions of soils—that is, species richness (alpha diversity, measured as amplicon sequence variants), community dissimilarity and ecosystem services—peaked in contrasting regions of the planet, and were associated with different environmental factors. Temperate ecosystems showed the highest species richness, whereas community dissimilarity peaked in the tropics, and colder high-latitude ecosystems were identified as hotspots of ecosystem services. These findings highlight the complexities that are involved in simultaneously protecting multiple ecological dimensions of soil. We further show that most of these hotspots are not adequately covered by protected areas (more than 70%), and are vulnerable in the context of several scenarios of global change. Our global estimation of priorities for soil nature conservation highlights the importance of accounting for the multidimensionality of soil biodiversity and ecosystem services to conserve soils for future generations.

Soils are essential to support terrestrial life on the planet¹. They are home to diverse assemblages of organisms across all major lineages of life from bacteria to invertebrates, and provide many different ecosystem services, such as soil fertility, carbon storage, waste decomposition, pest control and water retention^{3–5}, that are critical for food production and human well-being^{6–8}. However, soils are also highly vulnerable to anthropogenic disturbances such as climate change^{9,10} and land-use intensification (for example, land-use change, pollution and erosion^{11,12}). For an adequate conservation of soils, it is critical to consider and protect the multiple ecological dimensions supported by soils—from their biodiversity to the different ecosystem services they support. A first step in this direction is identifying the global ecological hotspots for soil nature conservation¹³ to inform and guide policymakers and conservation managers on how to extend nature conservation to the world belowground. Concurrently, establishing and negotiating global nature conservation policies and priorities (for example, the 2030 biodiversity targets¹⁴) requires knowledge about

the distribution of global biodiversity, including identifying ecological hotspots². Although these ecological hotspots were established decades ago for plants and animals, the corresponding information for soil biodiversity and ecosystem services does not exist, and is therefore absent from current assessments of biodiversity¹⁵. Developments in ecological modelling and soil macroecology have improved our understanding of the global distribution of multiple soil communities^{16–21} and their potential future trends^{9,22}. These studies have found possible mismatches between below- and aboveground biodiversity, which suggests that hotspots of plant diversity are poor proxies of belowground diversity²³ and, therefore, are unlikely to provide sufficient protection for life belowground. Although this may be true, plant species richness, such as that found in the tropics, is known to increase the diversity in soil organic matter compounds and, therefore, provide resources for a diverse soil microbiological community²⁴. This also opens the question of whether the hotspots of different soil ecological dimensions (for example, diversity, community composition

A list of affiliations appears at the end of the paper.

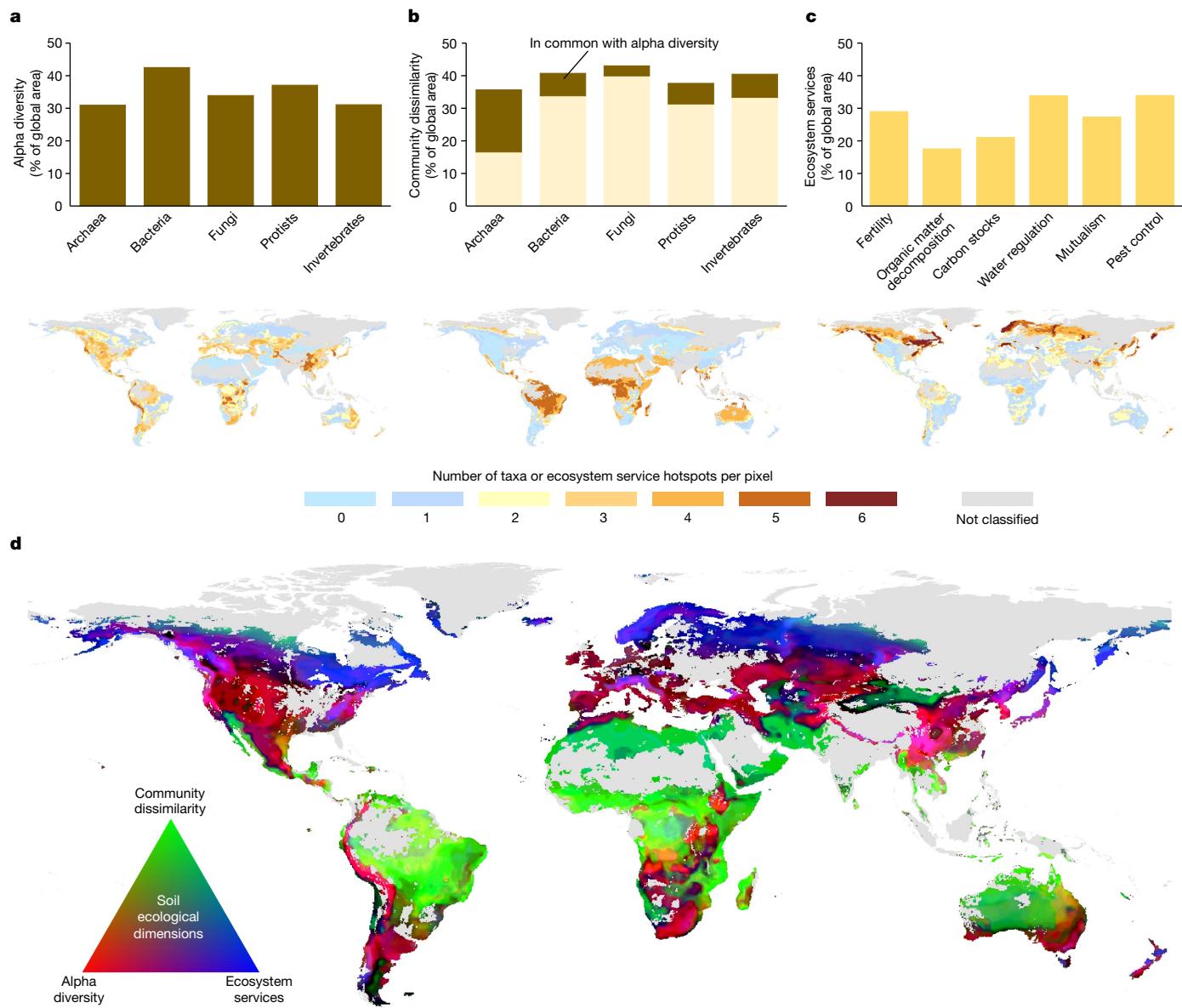


Fig. 1 | Current distribution of global soil ecological hotspots. **a–c**, Proportion of land occupied by hotspots of alpha diversity (**a**), community dissimilarity (**b**) and ecosystem services (**c**) (Supplementary Table 2 and Extended Data Figs. 4 and 5). The top row corresponds to the proportion of global area occupied by single taxa (**a,b**) or ecosystem services (**c**), and the bottom row to the global representation of accumulated hotspots across taxa (**a,b**) or ecosystem services (**c**). **d**, Together, these three ecological dimensions of soil create a soil nature

conservation profile in which both areas that maximize a given dimension and areas that allow for preserving a combination of global soil biodiversity hotspots are identified. Grey areas correspond to areas that were not assessed during the calculations owing to high uncertainty and insufficient environmental coverage (corresponding to 38.4% of the terrestrial world; Supplementary Fig. 5). A further estimation of spatial uncertainty for each dimension considered is provided in Supplementary Figs. 12 and 13.

and functions) coincide in space and how these are affected by global change, with studies pointing to diverging global patterns²². However, many recent developments are based on merged meta-analytical data, which are rarely measured using the same methods or do not simultaneously consider multiple soil ecological dimensions in the same locations. Unfortunately, we still lack globally standardized field surveys that explicitly consider the ecological multidimensionality of soils and which capture information on multiple soil taxonomic groups and ecosystem services simultaneously, across a wide range of global environmental conditions²⁵. Closing these knowledge gaps is essential to inform the establishment of nature conservation areas, steer management decisions and set effective policy targets that address the ecological conservation of soils.

Here, we combined machine learning models with a standardized global field survey, including 615 composite samples of topsoil from all continents and climates (Supplementary Fig. 1) to estimate the extent, associated environmental factors and vulnerabilities to climate change of the global hotspots of soil biodiversity and ecosystem services. Our dataset is based on more than 11,000 individual standardized observations and includes information on 16 attributes of biodiversity and ecosystem services (Methods and Supplementary Table 1). Our study moves beyond the analysis of alpha diversity (here based on soil DNA amplicon sequence variants; ASVs) and extends its scope to the community dissimilarity (that is, composition heterogeneity, based on Jaccard distance from presence-absence data) of five soil groups of organisms (archaea, bacteria, fungi, protists and invertebrates). Measuring the

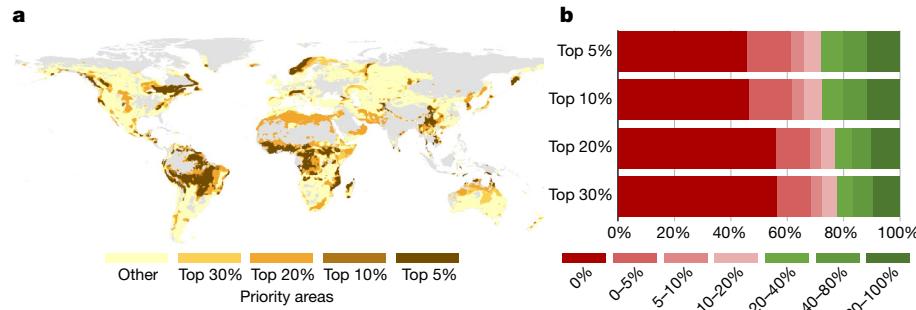


Fig. 2 | Current distribution of the priority areas for global soil nature conservation. **a**, Global soil nature conservation priorities. Spatial representation of the top 5%, 10%, 20% and 30% of areas with the highest accumulation of soil biodiversity and ecosystem services hotspots.

b, Proportion of those areas that are under some form of nature conservation regime. Grey areas correspond to areas that were not assessed during calculation.

hotspots of community dissimilarity and comparing them to those of alpha diversity allows us to identify areas with high local diversity that, at the same time, contain unique communities. In addition, to fully grasp the conservation potential of soil systems, soil functional properties related to six key soil ecosystem services were assessed: soil carbon storage (total soil organic carbon); fertility (total nitrogen, phosphorus, potassium and magnesium contents); terrestrial ecosystems in this study were not fertilized, and therefore, total nitrogen represents the stocks of nitrogen in soil organic matter); organic matter decomposition (three enzymes associated with starch and chitin degradation, and phosphorus mineralization); water retention (water-holding capacity); pest control (inverse of the proportion of soil-borne fungal phytopathogens); and mutualism (plant–mycorrhizal mutualism, an ecological relationship between plants and fungi that is beneficial to both partners, assessed as the proportion of mycorrhizal fungi). We acknowledge that our study does not cover the entire environmental spectra found on Earth, but it represents a large portion of the planet's environmental variability (Supplementary Fig. 1). Locations showing environmental conditions that were underrepresented in our study were excluded from our spatial analyses (Figs. 1–3).

Our analyses revealed that the assessed soil biodiversity and ecosystem service variables are associated with contrasting environmental factors at the global scale (Extended Data Figs. 1 and 2). For example, although soil pH was the main factor associated with the alpha diversity of soil fungi and bacteria, soil organic matter (soil carbon and nitrogen contents) was positively associated with the alpha diversity of protists and invertebrates (Extended Data Figs. 2 and 3), and elevation was positively correlated with the alpha diversity of archaea. In the case of the assessed ecosystem services, soil pH was positively associated with soil carbon content, water retention and pest control, whereas temperature was associated with organic matter decomposition and fertility, and precipitation seasonality with mutualism (Extended Data Figs. 2 and 3). Although many of these environmental associations are well-described in the literature^{26,27}, the fact that different soil ecological dimensions could be predicted by contrasting environmental factors was much less clear, owing to the lack of standardized global field surveys. These contrasting associations and environmental drivers explain the different global distributions that were found for each ecological dimension and reveal that important trade-offs may exist when considering the nature conservation of multi-faceted soil systems. To further visualize these trade-offs, we used machine learning models based on random forest spatial regression, together with available current data and future projections for both climate and land-use change (2015–2070), to predict the distribution of soil biodiversity and ecosystem services and assess their major drivers according to several future scenarios (shared socio-economic pathways: SSP1, global sustainability; SSP3, regional rivalry; SSP4, inequality; and SSP5, fossil-fuelled development²⁸). We

standardized each of these spatial distributions and used a Getis-Ord Gi* spatial clustering algorithm to obtain a representation of the global hotspots (clusters of statistically high values) for the modelled distribution of each single biodiversity and ecosystem service variable. These were then aggregated into each soil ecological dimension (Fig. 1a–c). To further strengthen our conclusions, we performed a comparison across multiple methods (Supplementary Figs. 7 and 10) and an uncertainty assessment for the spatial predictions (Supplementary Figs. 5 and 11–13). A rationale supporting the spatial analysis from our standardized survey, and explaining the limitations of our approach, is available in the Methods.

Furthermore, we found that different ecological dimensions for soil conservation peak in different regions of the Earth (Fig. 1a–c). Model fitness (measured as overall training R^2) varied between 0.855 and 0.914 for alpha diversity and community dissimilarity (Supplementary Table 5), and between 0.801 and 0.936 for ecosystem services (Supplementary Table 6). Hotspots of alpha diversity tend to have a wider distribution across the world, peaking in temperate and Mediterranean regions, as well as in alpine tundra (overall occupying between 30.9% (for archaea) and 42.4% (for bacteria) of the world). However, hotspots of community dissimilarity occur around two contrasting global conditions—tropical systems and drylands (overall occupying between 35.7% (for archaea) and 43.0% (for fungi) of the world). For fungi, our results were further compared and validated with an independent dataset (Supplementary Fig. 6). Although a higher alpha diversity may intuitively imply a direct decrease in dissimilarity, directly varying in tandem, our results show that at the global scale this is not the case (Supplementary Table 15). Archaea exhibited the highest proportion of shared hotspot areas (19%), with all other groups showing a proportion of less than 8% (alpha diversity and community dissimilarity for the same taxa; Supplementary Table 3). Our findings further suggest the existence of important trade-offs in soil nature conservation priorities (Extended Data Fig. 3). For example, locations with a higher alpha diversity tend to be less dissimilar, and only a small proportion of locations were found to support both high dissimilarity and high alpha diversity (Fig. 1b). This proportion is smaller for fungi (3.9%) and larger for archaea (19.0%; Supplementary Table 2). Similarly, locations with a higher dissimilarity tend to have a lower soil carbon content, lower fertility and a higher proportion of plant pathogens (Extended Data Fig. 3). Moreover, our global maps indicate that alpha diversity (Fig. 1a), community dissimilarity (Fig. 1b) and ecosystem services (Fig. 1c) have their hotspots in mostly contrasting regions of the planet, existing only in a few locations that support high levels of more than one of these dimensions (0.1% of the evaluated areas in the world, based on Fig. 1). This contrasts with results found for other biodiversity groups such as plants and mammals^{29,30} and supports previous findings of a mismatch between soil biodiversity and other taxonomic groups²³.

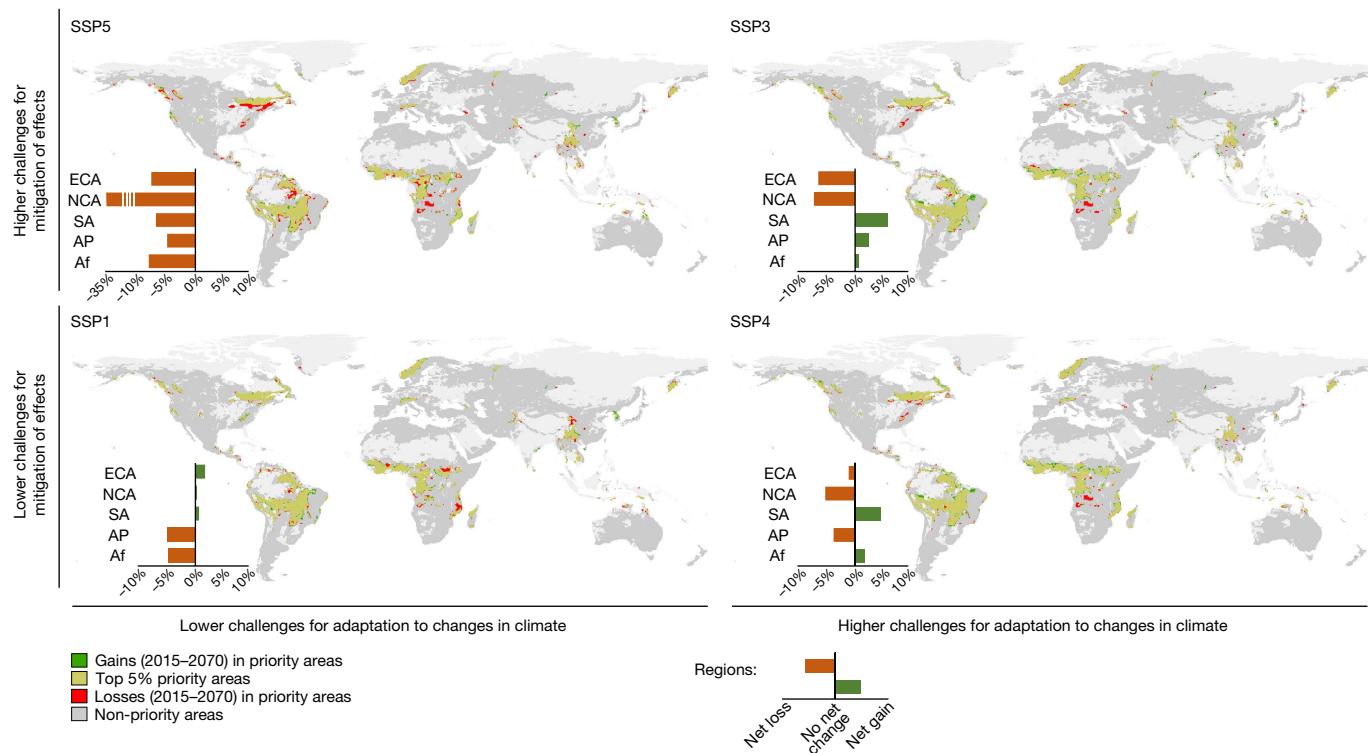


Fig. 3 | Predicted changes in the total area of soil nature conservation priorities under four different future shared socio-economic pathways.

Graphs and spatial representations of the expected changes (gains and losses between 2015 and 2070) in the total area of soil nature conservation priorities (top 5%) according to four different shared socio-economic pathways⁸:

SSP1 (global sustainability); SSP3 (regional rivalry); SSP4 (inequality); and SSP5 (fossil-fuelled development). Af, Africa; AP, Asia Pacific; ECA, Europe and Central Asia; NCA, North America and Central America; SA, South America. Light grey areas correspond to areas that were not assessed during the calculations.

(Supplementary Table 13.) Although globally, tropical and arid systems were mostly classified as locations with a relatively low alpha diversity across taxa, these areas are hotspots for soil community dissimilarity, and support the most unique soil community assemblies (Fig. 1b). In the case of bacteria, for example, locations with a high pH support a higher richness (for example, in temperate systems); however, these are always similar organisms that thrive in neutral–alkaline soils. This suggests that, although local diversity in tropical systems may be low (for example, as a consequence of acidic soils), these environmentally contrasting areas of the globe may contain unique communities, which in turn may result in a high gamma (regional) diversity. Although this hypothesis has already been suggested^{16,22}, our study represents—to our knowledge—the first robust confirmation of it.

Our results highlight the fact that preserving soils from a nature conservation perspective requires a comprehensive approach that considers multiple soil ecological dimensions such as alpha diversity, community dissimilarity and ecosystem services in the context of a nature conservation profile (Fig. 1d). Being able to position a given area within this soil nature conservation profile allows conservation goals to be set that effectively target the preservation of soil communities and their effects on ecosystems³¹. For example, an area that falls into a community dissimilarity hotspot may focus on indicators and conservation goals to track and prevent species losses, because these may be unique to its limits, whereas an area in an ecosystem service hotspot may favour indicators that target ecosystem service supply. This does not imply that conservation areas should not prioritize all soil ecological dimensions, but rather that management strategies and conservation targets should be adjusted to the ecological reality of each region and conservation area². Moreover, although local approaches are still needed to refine the local distribution of these hotspots³², these results also suggest that no particular region can protect all dimensions

of soil conservation, making a further argument for global cooperation and for establishing global targets for soil nature conservation.

Given the contrasting regions that support the highest biodiversity and services, identifying which ecological dimension is the most relevant for the conservation of soil ecological conditions is not a simple task. Some ecosystems depend on a high alpha diversity, whereas others do not need such high levels of alpha diversity to properly function, but rely on more dissimilar soil communities³³ (Fig. 1d). Conversely, although for microbial communities³⁴, functional redundancy driven by community composition may be more important than alpha diversity per se, in general terms, ecosystems with a lower alpha diversity are likely to be more sensitive to ecosystem change³⁵—and for that, more targeted conservation actions are required. Although it is known that soil organisms have a crucial role in the supply of ecosystem services³⁶, it is not clear whether it is biodiversity per se that governs this entire process. For example, some specific ecosystem services may depend on the presence of only a few species, such as specific components of the soil nitrogen cycle³⁷, whereas others are the result of the activity of many species with high levels of redundancy, such as soil respiration³⁸. Therefore, we defined priority areas for soil nature conservation as areas that support relatively high levels of either soil biodiversity or ecosystem services. We were then able to identify key regions of the planet that surpass high thresholds of either biodiversity or ecosystem service provision—the hotspots for soil nature conservation (Fig. 2). Considering the areas with the highest accumulation of soil biodiversity hotspots (top 5% of areas), it is possible to identify tropical systems and substantial areas in North America, northern Europe and Asia as having high priority for nature conservation (Fig. 2a). These areas maximize different dimensions of soil ecology and may thus require integrative strategies, not only from a nature conservation perspective but also considering the socio-economic appropriation of belowground

systems. In this regard, it is notable that around 50% of these priority areas for global nature conservation are not under any form of nature conservation, and that only around 10% correspond to areas that are fully protected (Fig. 2b). As these priority areas are the areas with the highest relevance for nature conservation, and given that at present, soils do not have any specific nature conservation targets, this is a worrisome state for the conservation of soil biodiversity worldwide². This situation is also observed if we consider other thresholds for the soil nature conservation potential (Fig. 2b). Although soil conservation may not be able to maximize all ecological dimensions at the same time, and each region may have different specificities with targeted research required, a number of actions may be considered. These include nature-based solutions in land management for enhancing ecosystem services³⁹; landscape-level actions such as the preservation of permanent forest and natural coverage in the surrounding of managed systems⁴⁰; or nature-based solutions that are focused on restoring or improving soil functional outputs⁴¹. Our work provides key information for regional and continental decision-makers to develop goals for nature protection that specifically target soil systems and biodiversity, including identifying areas with a high potential to establish soil-based nature conservation areas.

In the context of climate and land-use change, nature conservation areas and targets will need to adapt to new conditions and also focus on mitigating potential impacts^{42–44}. Thus, focusing on the priority areas for global soil nature conservation (top 5%), we conducted an additional analysis to predict the future changes in hotspots according to four shared socio-economic pathways (Fig. 3; 2015–2070). Our projections highlight the fact that the soil nature conservation hotspots will change as a result of climate and land-use change linked to substantial declines in both alpha diversity and ecosystem services. Globally, across scenarios, net differences between 2015 and 2070 range from 1.5% net gains in SSP3 and –12.2% net losses in SSP5. In most cases, these net changes actually hide substantial losses of present soil nature conservation priority areas, with 7.1% (SSP4) to 17.5% (SSP5) of current areas being lost globally across different future scenarios (Supplementary Table 7). Our results reveal that most of the net area losses are related to declines in ecosystem services, particularly carbon stocks (average loss across scenarios = –6.8%), mutualism (–3.8%) and litter decomposition (–3.6%), and in the alpha diversity of specific groups, particularly invertebrates (–2.6%), fungi (–1.3%) and archaea (–1.1%). Our projections also show that new regions will emerge as key areas for soil nature conservation across the world, corresponding to expansions that range from 5.3% in SSP5 (relative to the current area) to 9.5% in SSP3. Notably, scenarios that consider higher challenges for adaptation to climate change motivated by higher regional income inequality and rivalry (SSP3 and SSP4)⁴⁵ also show the most positive effects for maintaining or expanding current nature conservation priorities for soils, particularly in Africa and South America (Fig. 3). Overall, these positive effects are mostly expected in the global south, with systematic negative effects in the global north across scenarios. In fact, the only scenario in which the global north has slight net gains (0.3–2.0%) corresponds to the so-called 'sustainability scenario' (SSP1). Nevertheless, in this scenario, most of the rest of the world shows substantial net losses (–5.7% in Africa and –5.9% in Asia Pacific) or only mild net gains (0.8% in South America), owing to expected increases in global economic development (Fig. 3). This is even more concerning when we consider reports that show that 30% of the population across tropical countries is highly dependent on nature⁴⁶. Across all regions, the fossil-fuelled economy scenario (SSP5) produces the strongest net losses, with regions losing priority areas from –5.9% for Asia Pacific to –31.8% for North America (most of these losses are driven by decreases in ecosystem services). Furthermore, our results suggest that the present simplistic view on carbon-based targets provides little protection for all soil ecological dimensions. In fact, the sustainability scenario shows an overall global improvement in ecosystem services, with soil carbon leading these improvements,

but with clear losses in alpha diversity. Together, these results indicate that hotspots of soil biodiversity and ecosystem services are highly threatened by future climatic and land-use changes, and stress the need for immediate protection of these locations. Our findings also suggest that these hotspots might move in the future, with current sanctuaries of soil biodiversity being subject to degradation.

In summary, based on a large global standardized survey, including 16 biodiversity and ecosystem service variables, our work provides an estimate of the global hotspots for the conservation of multiple ecological dimensions of soil. We identify critical unique areas for the conservation of soil biodiversity and ecosystem services at the global scale, with soil alpha diversity, dissimilarity and services peaking in temperate, tropical and boreal regions, respectively. Although recent literature highlights the need for extending nature conservation to ensure global sustainability and the preservation of biodiversity⁴⁷, it also underlines that this increased protection requires context-based solutions⁴⁸. By revealing important trade-offs in soil biodiversity and ecosystem services, we also show that no particular region of the world will be able to simultaneously protect all these ecological dimensions. Therefore, the conservation of soil biodiversity and ecosystem services requires an integrated approach that probably should not focus on locally maximizing all ecological dimensions at the same time. In addition, the fact that we found that these three ecological dimensions do not necessarily match in terms of their spatial hotspots also shows the complexity of soil ecosystems and emphasizes the difficulties that land managers and policymakers face when designing soil conservation measures. Nevertheless, we also show that these nature conservation priority areas are at present highly unprotected, with less than 10% of these locations being under adequate conservation status. We acknowledge that our study is only a first step towards understanding and mapping the global hotspots for soil nature conservation, and that high-resolution monitoring systems and multiple time periods are needed to better guide regional conservation and policy options². But our work suggests that current priority areas for soil nature conservation are vulnerable to drivers of global change in all future scenarios considered, and stresses the need for immediate nature conservation targeting and protection of these regions. This information should enable governments and decision-makers to set soil nature conservation as a priority in the context of the negotiations of the 2030 biodiversity targets, thus paving the way for a more integrative view of nature.

Online content

Any methods, additional references, Nature Research 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/s41586-022-05292-x>.

1. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
2. Guerra, C. A. et al. Tracking, targeting, and conserving soil biodiversity. *Science* **371**, 239–241 (2021).
3. Wall, D. H. et al. (eds) *Soil Ecology and Ecosystem Services* (Oxford University Press, 2012).
4. Jansson, J. K. & Hofmockel, K. S. Soil microbiomes and climate change. *Nat. Rev. Microbiol.* **18**, 35–46 (2020).
5. de Vries, F. T. et al. Soil food web properties explain ecosystem services across European land use systems. *Proc. Natl. Acad. Sci. USA* **110**, 14296–14301 (2013).
6. Adhikari, K. & Hartemink, A. E. Linking soils to ecosystem services—a global review. *Geoderma* **262**, 101–111 (2016).
7. Pereira, P., Bogunovic, I., Muñoz-Rojas, M. & Brevik, E. C. Soil ecosystem services, sustainability, valuation and management. *Curr. Opin. Environ. Sci. Health* **5**, 7–13 (2018).
8. Wall, D. H., Nielsen, U. N. & Six, J. Soil biodiversity and human health. *Nature* **528**, 69–76 (2015).
9. Delgado-Baquerizo, M. et al. The proportion of soil-borne pathogens increases with warming at the global scale. *Nat. Clim. Chang.* **10**, 550–554 (2020).
10. Rillig, M. C. et al. The role of multiple global change factors in driving soil functions and microbial biodiversity. *Science* **366**, 886–890 (2019).

11. Guerra, C. A. et al. Global vulnerability of soil ecosystems to erosion. *Landsc. Ecol.* **35**, 823–842 (2020).
12. Geisen, S., Wall, D. H. & van der Putten, W. H. Challenges and opportunities for soil biodiversity in the Anthropocene. *Curr. Biol.* **29**, R1036–R1044 (2019).
13. Jung, M. et al. Areas of global importance for conserving terrestrial biodiversity, carbon and water. *Nat. Ecol. Evol.* **5**, 1499–1509 (2021).
14. Xu, H. et al. Ensuring effective implementation of the post-2020 global biodiversity targets. *Nat. Ecol. Evol.* **5**, 411–418 (2021).
15. Diaz, S. et al. (eds). *Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES, 2019);* <https://zenodo.org/record/3553579#.YyhlXbMK70>
16. Phillips, H. R. P. et al. Global distribution of earthworm diversity. *Science* **366**, 480–485 (2019).
17. van den Hoogen, J. et al. Soil nematode abundance and functional group composition at a global scale. *Nature* **572**, 194–198 (2019).
18. Delgado-baquerizo, M. et al. A global atlas of the dominant bacteria found in soil. *Science* **325**, 320–325 (2018).
19. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
20. Xu, X., Thornton, P. E. & Post, W. M. A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems: global soil microbial biomass C, N and P. *Glob. Ecol. Biogeogr.* **22**, 737–749 (2013).
21. Djukic, I. et al. Early stage litter decomposition across biomes. *Sci. Total Environ.* **628–629**, 1369–1394 (2018).
22. Guerra, C. A. et al. Global projections of the soil microbiome in the Anthropocene. *Glob. Ecol. Biogeogr.* **30**, 987–999 (2021).
23. Cameron, E. K. et al. Global mismatches in aboveground and belowground biodiversity. *Conserv. Biol.* **33**, 1187–1192 (2019).
24. El Moujahid, L. et al. Effect of plant diversity on the diversity of soil organic compounds. *PLoS One* **12**, e0170494 (2017).
25. Guerra, C. A. et al. Blind spots in global soil biodiversity and ecosystem function research. *Nat. Commun.* **11**, 3870 (2020).
26. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **103**, 626–631 (2006).
27. Tedersoo, L. et al. Regional-scale in-depth analysis of soil fungal diversity reveals strong pH and plant species effects in Northern Europe. *Front. Microbiol.* **11**, 1953 (2020).
28. Popp, A. et al. Land-use futures in the shared socio-economic pathways. *Glob. Environ. Change* **42**, 331–345 (2017).
29. Dornelas, M. et al. Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296–299 (2014).
30. Ego, B., Reyers, B., Rouget, M., Bode, M. & Richardson, D. M. Spatial congruence between biodiversity and ecosystem services in South Africa. *Biol. Conserv.* **142**, 553–562 (2009).
31. Jürgens, N. et al. The BIOTA Biodiversity Observatories in Africa—a standardized framework for large-scale environmental monitoring. *Environ. Monit. Assess.* **184**, 655–678 (2012).
32. Wyborn, C. & Evans, M. C. Conservation needs to break free from global priority mapping. *Nat. Ecol. Evol.* **5**, 1322–1324 (2021).
33. Hautier, Y. et al. Local loss and spatial homogenization of plant diversity reduce ecosystem multifunctionality. *Nat. Ecol. Evol.* **2**, 50–56 (2018).
34. Zhou, Z., Wang, C. & Luo, Y. Meta-analysis of the impacts of global change factors on soil microbial diversity and functionality. *Nat. Commun.* **11**, 3072 (2020).
35. Eisenhauer, N., Schulz, W., Scheu, S. & Jousset, A. Niche dimensionality links biodiversity and invasibility of microbial communities. *Funct. Ecol.* **27**, 282–288 (2013).
36. Wagg, C., Bender, S. F., Widmer, F. & van der Heijden, M. G. A. Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proc. Natl Acad. Sci. USA* **111**, 5266–5270 (2014).
37. Haines-Young, R. H. & Potschin, M. B. in *Ecosystems Ecology: A New Synthesis* (eds Raffaelli, D. G. & Frid, C. L. J.) Ch. 6 (2012).
38. Smith, L. C. et al. Large-scale drivers of relationships between soil microbial properties and organic carbon across Europe. *Glob. Ecol. Biogeogr.* **30**, 2070–2083 (2021).
39. Keesstra, S. et al. The superior effect of nature based solutions in land management for enhancing ecosystem services. *Sci. Total Environ.* **610–611**, 997–1009 (2018).
40. Le Provost, G. et al. Contrasting responses of above- and belowground diversity to multiple components of land-use intensity. *Nat. Commun.* **12**, 3918 (2021).
41. Tanneberger, F. et al. The power of nature-based solutions: how peatlands can help us to achieve key EU sustainability objectives. *Adv. Sustain. Syst.* **5**, 2000146 (2021).
42. Johnston, A. et al. Observed and predicted effects of climate change on species abundance in protected areas. *Nat. Clim. Chang.* **3**, 1055–1061 (2013).
43. Hannah, L. et al. Protected area needs in a changing climate. *Front. Ecol. Environ.* **5**, 131–138 (2007).
44. Gallardo, B. et al. Protected areas offer refuge from invasive species spreading under climate change. *Glob. Chang. Biol.* **23**, 5331–5343 (2017).
45. O’Neill, B. C. et al. The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century. *Glob. Environ. Change* **42**, 169–180 (2017).
46. Fedele, G., Donatti, C. I., Bornacelly, I. & Hole, D. G. Nature-dependent people: mapping human direct use of nature for basic needs across the tropics. *Glob. Environ. Change* **71**, 102368 (2021).
47. Visconti, P. et al. Protected area targets post-2020. *Science* **364**, 239–241 (2019).
48. Allan, J. R. et al. The minimum land area requiring conservation attention to safeguard biodiversity. *Science* **376**, 1094–1101 (2022).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

¹German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. ²Institute of Biology, Martin Luther University Halle Wittenberg, Halle(Saale), Germany. ³Institute of Biology, Leipzig University, Leipzig, Germany. ⁴Institute of Integrative Biology, Department of Environment Systems Science, ETH Zürich, Zürich, Switzerland. ⁵Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, Australia. ⁶Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales, Australia. ⁷Global Centre for Land-Based Innovation, Western Sydney University, Penrith, New South Wales, Australia. ⁸Institute of Grassland Science, School of Life Science, Northeast Normal University, Key Laboratory of Vegetation Ecology of the Ministry of Education, Jilin Songnen Grassland Ecosystem National Observation and Research Station, Changchun, China. ⁹Departamento de Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, Seville, Spain. ¹⁰GEMA Center for Genomics, Ecology and Environment, Faculty of Interdisciplinary Studies, Universidad Mayor, Huechuraba, Chile. ¹¹Instituto de Ecología & Biodiversidad (IEB), Santiago, Chile. ¹²Natural History Museum, Obafemi Awolowo University, Ille-Ife, Nigeria. ¹³CEBAS-CSIC, Campus Universitario de Espinardo, Murcia, Spain. ¹⁴Department of Plant Biology and Ecology, University of Seville, Seville, Spain. ¹⁵Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid, Spain. ¹⁶Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal. ¹⁷Misión Biológica de Galicia, Consejo Superior de Investigaciones Científicas, Pontevedra, Spain. ¹⁸Slovenian Forestry Institute, Ljubljana, Slovenia. ¹⁹Department of Entomology, College of Agricultural, Human, and Natural Resource Sciences, Washington State University, Pullman, WA, USA. ²⁰College of Resources and Environment, Huazhong Agricultural University, Wuhan, China. ²¹Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa. ²²Department of Soil Science, College of Agriculture and Bioresources, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. ²³Laboratorio de Ecología Integrativa, Instituto de Ciencias Biológicas, Universidad de Talca, Talca, Chile. ²⁴CEAZA, Universidad Católica del Norte, Coquimbo, Chile. ²⁵Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, India. ²⁶Departamento de Botánica, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. ²⁷Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, Madrid, Spain. ²⁸Departamento de Biología, Instituto Universitario de Investigación Marina (INMAR), Universidad de Cádiz, Puerto Real, Spain. ²⁹Mycology and Microbiology Center, University of Tartu, Tartu, Estonia. ³⁰College of Science, King Saud University, Riyadh, Saudi Arabia. ³¹Departamento de Botánica e Ecología, Instituto de Biociencias, Universidade Federal de Mato Grosso, Cuiabá, Brazil. ³²Grupo de Investigación en Biodiversidad y Cambio Global (GIBCG), Departamento de Ciencias Básicas, Universidad del Bío-Bío, Chillán, Chile. ³³Microbiome Network and Department of Agricultural Biology, Colorado State University, Fort Collins, CO, USA. ³⁴Department of Natural Resources, Agricultural Research Organization, Institute of Plant Sciences, Gilat Research Center, Negev, Israel. ³⁵State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China. ³⁶Laboratorio de Biodiversidad y Funcionamiento Ecosistémico, Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, Seville, Spain. ³⁷Unidad Asociada CSIC-UPO (BioFun), Universidad Pablo de Olavide, Seville, Spain. [✉]e-mail: carlos.guerra@idiv.de; m.delgado.baquerizo@csic.es

Methods

A global standardized survey to investigate topsoil biodiversity and function

We used composite topsoil samples from global field surveys that were conducted between 2016 and 2019 following standardized field protocols. This global field survey includes 151 locations from all continents and 23 countries, from which 615 composite topsoil samples were collected, providing a large representation of all climatic and vegetation biomes in the planet (Supplementary Fig. 1). The locations of the soil samples were not established following a random protocol but rather were selected taking into account the local representativeness of the vegetation within the ecosystem types sampled. In global terms, the approach aimed to include as much climatic and edaphic variability as possible given the constraints of such a sampling scheme. Between three and five composite soil (top around 0–10 cm) samples (from five–ten soil cores) were collected in these locations (ranging from 0.09–0.25 ha) following the protocol described in a previous study⁴⁹. By including multiple composite samples within each location, we aimed to account for within-location heterogeneity variation in soil properties, biodiversity and services. We focused on the topsoils, because they are known to hold the largest proportion of soil biodiversity, and constitute the critical zone that supports key soil processes from organic matter decomposition to plant–soil interactions. A portion of these soils was frozen (-20°C) after sampling for molecular analyses, and another portion was air-dried and used for determining soil properties. We recognize that although our dataset provides a fairly complete coverage of global environmental conditions, an increase of sampling locations in less-represented regions of the globe would increase the strength of the study. To this respect, we aimed at adequately representing the spatial limitations of our study by eliminating and masking out all the regions that were poorly represented (Supplementary Fig. 5). Reaching this spatial representation was not easy, owing to logistic limitations (for example, the absence of local resources for sample preservation and consequent material degradation²⁵), as well as war and transport embargos. These issues disproportionately affect these underrepresented regions and result in notable gaps in Africa and Southeast Asia. The dataset is available at <https://doi.org/10.6084/m9.figshare.20221713>.

Soil biodiversity

The alpha diversity (corresponding to the number of phylotypes) and community dissimilarity (averaged Jaccard distance across samples from presence–absence matrices to account for dissimilarity in phylotypes, measured as ASVs, rather than in their proportions) of archaea, bacteria, fungi, protists and invertebrates was determined using amplicon sequencing technology (Illumina MiSeq platform) following a previously published protocol⁵⁰. Both of these measurements are crucial to understand the nature and conservation potential of specific areas. Whereas alpha diversity refers to the number of species (or ASVs in this case) contained in a particular location, typically seen as priority areas for nature conservation, community dissimilarity refers to the uniqueness of the community, indicating the presence of specific species that are not common elsewhere. The latter also represents an important aspect for the selection of new conservation areas⁵¹. Soil DNA was extracted using the Powersoil DNA Isolation Kit (MoBio Laboratories) according to the manufacturer's instructions. A portion of the bacterial and archaeal 16S and eukaryotic 18S rRNA genes were sequenced using the 515F/806R and Euk1391f/EukBr primer sets^{52–54}, respectively. Bioinformatics processing was performed using a combination of QIIME⁵⁵, USEARCH⁵⁶ and UNOISE3^{57,58}. Phylotypes (ASVs) were identified at the 100% identity level. The ASV abundance tables were rarefied at 5,000 (bacteria by 16S rRNA gene), 100 (archaea by 16S rRNA gene), 2,000 (fungi by 18S rRNA gene), 1,000 (protists by 18S rRNA gene), and 250 (invertebrates by 18S rRNA gene) sequences

per sample, respectively, to ensure an even sampling depth within each belowground group of organisms. Protists are defined as all eukaryotic taxa, except fungi, invertebrates (Metazoa) and vascular plants (Stratophyta). Note that not all samples passed our rarefaction cut-off. The total number of samples included in each soil group is provided in Supplementary Table 2.

Library preparation and sequencing. Triplicate PCR reactions were performed for each of the extracted DNA samples, and we included and sequenced multiple negative controls per plate to check for possible contamination. Each 25- μl PCR reaction contained: 12.5 μl Promega GoTaq Hot Start Colorless Master Mix; 0.5 μl of each bar-coded primer (bacterial 16S, 515F (5'-GTGCCAGCMGCCGCGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') or eukaryotic 18S, Euk1391f (5'-GTACACCGCCGTC-3') and EukBr (5'-TGATCCTTCTGCAGG TTCACCTAC-3'); 10.5 μl water; 1 μl of template DNA. 'Fusion' primers also included Illumina adapters and 12-bp barcodes to enable multiplexed sequencing. PCR conditions for bacterial 16S rDNA amplifications were 94 °C for 3 min; 35 cycles of 94 °C for 45 s, 50 °C for 60 s, 72 °C for 90 s; 72 °C for 5 min. PCR conditions for eukaryotic 18S rDNA amplifications were 94 °C for 3 min; 35 cycles of 94 °C for 45 s, 57 °C for 60 s, 72 °C for 90 s; 72 °C for 10 min. PCR products were cleaned with the MoBio Ultra Clean PCR Clean-Up Kit. Next, we performed PCR-mediated Nextera barcode ligation following the manufacturer's instructions, adding unique barcodes onto amplicons, to allow for multiplexed sequencing. Samples were normalized with the Sequal-PREP Normalization Plate Kit (Invitrogen) before sequencing on the Illumina MiSeq platform.

DNA was first cleaned up using AMPure Xp beads (Beckman Coulter) and then quantified using the automated fluorescence-based PicoGreen assay (Invitrogen). The cleaned DNA was normalized to 1.5 ng μl^{-1} and a total of 7 ng of the input DNA was used for each amplicon PCR reaction. Illumina's instructions do not recommend pooled three PCRs (which should not be considered as technical replicates), and one PCR reaction was performed per amplicon. To minimize the PCR bias in the sequencing, the number of PCR cycles was reduced to 25. In detail, the PCR conditions for bacterial 16S rDNA amplification were 95 °C for 3 min, 25 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, 72 °C for 5 min, then hold at 4 °C; and the PCR conditions for eukaryotic 18S rDNA amplification were: 94 °C for 5 min, 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, 72 °C for 5 min, then hold at 4 °C. The Illumina forward overhang (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3') and reverse overhang (5'-GTCTCGTGGCTGGAGATGTGTATAAGAGACAG-3') adapters were included in the amplicon PCR. Each sample was barcoded with two eight-base indices using the Illumina Nextera XT Index Kit. The 15- μl reaction system was prepared with KAPA HotStart ReadyMix Kit (Merck KGaA). A ZymoBIOMICS Microbial Community DNA Standard was used as a positive control together with the samples to assess the bias in PCR. Molecular analysis of the full-length ITS region for fungi was performed using ITS9mun/ITS4ngsUni primer sets and PacBio third-generation sequencing as described previously⁵⁹. Our sequencing run yielded 3861042 fungal sequences (18S), 534286 invertebrate sequences (18S), 2388020 protist sequences (18S), 8901512 bacterial sequences (16S) and 168749 archaeal sequences (16S).

Bioinformatics. Bioinformatic processing was performed using a combination of QIIME, USEARCH and UNOISE. In brief, data were demultiplexed and primers were trimmed before further analyses. Default parameters were followed in the USEARCH pipeline, except that the bases with a quality score lower than 20 were end-trimmed from the forward or reverse primer reads to minimize the mismatch in merging and to maximize the portion of successful mergers. A maximum of expected error was set to 1.0 for the merged reads quality filtering using USEARCH⁵⁶. zOTUs (or Amplicon Sequence Variant) were gained by denoising (error-correction) the dereplicated merged reads using

Article

UNOISE3 (ref. ⁵⁸). Representative sequences of ASVs were annotated against the Silva database⁶⁰ in QIIME⁵⁵ using UCLUST⁵⁶. The 18S taxonomy annotation used both SILVA and the Protist Ribosomal Reference database (PR², <https://pr2-database.org/>) (ref. ⁶⁰). The resulting ASV tables were rarefied at 5,000 (bacteria by 16S rRNA gene), 100 (archaea by 16S rRNA gene), 2,000 (fungi by 18S rRNA gene), 1,000 (protists by 18S rRNA gene) and 250 (invertebrates by 18S rRNA gene) sequences per sample, respectively, to ensure an even sampling depth within each belowground group of organisms. They were then imported into QIIME⁵⁵ for downstream analysis including diversity and community composition.

Rarefaction resolution and primer-set cross-validation

Rarefaction resolution. First, we performed additional analyses to provide evidence that our choice of rarefaction level did not affect our results or conclusions. Here, using the samples with the highest sequence per sample yield, we tested for the effect of different levels of rarefaction on soil biodiversity. We found highly statistically significant correlations between the diversities of soil archaea (rarefied at 100 versus 500 sequences per sample; Spearman $\rho = 0.764$; $P < 0.001$; $n = 128$), bacteria (rarefied at 5,000 versus 10,000 sequences per sample; Spearman $\rho = 0.992$; $P < 0.001$; $n = 509$), fungi (rarefied at 2,000 versus 10,000 sequences per sample; Spearman $\rho = 0.971$; $P < 0.001$; $n = 88$), protists (rarefied at 1,000 versus 5,000 sequences per sample; Spearman $\rho = 0.971$; $P < 0.001$; $n = 287$) and invertebrates (rarefied at 250 versus 1,000 sequences per sample; Spearman $\rho = 0.952$; $P < 0.001$; $n = 274$), for a subset of samples for which high numbers of sequences were available. These results are supported by previous independent global surveys providing evidence that rarefaction options do not influence global patterns in microbial communities^{50,61}. See rarefaction curves in Supplementary Fig. 3.

Primer-set cross-validation. Next, we provide additional evidence that primer sets are not influencing the global patterns reported here. For a subset of samples, we generated additional molecular information for fungal (ITS PacBio sequencing; ITS9mun/ITS4ngsUni primer sets) and bacterial (16S rRNA MiSeq Sequencing; 341F/805R primer sets) data. We found that the richness of soil microbial communities used in this study was highly significantly and positively correlated to those using these alternative primer sets both for bacteria (Spearman $\rho = 0.403$; $P < 0.001$; $n = 128$) and for fungi (Spearman $\rho = 0.656$; $P < 0.001$; $n = 228$). Similarly, the community compositions of bacteria (Spearman $\rho = 0.479$; $P < 0.001$; $n = 128$) and fungi (Spearman $\rho = 0.414$; $P < 0.001$; $n = 228$) were significantly and positively correlated to those using these alternative primer sets. We also found that the main predictors of bacterial and fungal richness (soil pH in both cases; Supplementary Fig. 2) in this study followed the same pattern for bacterial and fungal richness using alternative primer sets (Supplementary Fig. 4). The 18s primer sets used here to describe protists and invertebrates are the gold standard for the sequencing of these organisms^{54,62} and have been previously cross-validated in the literature^{63–65}. We acknowledge that there are multiple alternative primer sets, especially when specifically targeting particular groups of organisms within protists (for example, the mtDNA cytochrome oxidase I (COI) gene). Nevertheless, although specific primers may deliver higher resolution for specific groups, these are known to be inefficient in identifying a wide range of organisms from environmental samples⁶⁶.

Mapping the distribution of fungal functional guilds. Finally, to provide further evidence that 18S rRNA MiSeq sequencing can in this case provide a good representation of the global patterns in soil-borne mycorrhizal fungi and fungal potential plant pathogens, we compared the global patterns (see the section "Global hotspots of soil biodiversity and services") in the proportion of soil-borne mycorrhizal fungi and fungal potential plant pathogens determined using 18S rRNA MiSeq

sequencing with the subset of data including ITS PacBio sequencing (see the section "Primer-set cross-validation"). Our results showed that the proportion of soil-borne mycorrhizal fungi and fungal potential plant pathogens determined using two independent methods followed similar patterns and had a strong and positive correlation worldwide (Supplementary Fig. 6 and Supplementary Tables 8 and 9), thus allowing us to tentatively use the 18S rRNA gene as a proxy for phylotype richness. The number of arbuscular mycorrhizal fungi phylotypes retrieved from ITS PacBio sequencing was not enough to conduct this analysis, so we used ectomycorrhizal fungi in our mapping comparison.

Soil ecosystem services

Six soil functions directly related to key ecosystem services were determined using highly standardized methods: water retention (water-holding capacity), fertility (nitrogen, phosphorus, potassium and magnesium content), carbon storage (total soil organic carbon content), mutualism (proportion of arbuscular and ectomycorrhizal fungi), pest control (inverse of the proportion of soil-borne potential plant pathogens; as defined previously⁶⁷) and organic matter decomposition (three enzymes associated with the carbon, nitrogen and phosphorus cycle). Percentage of water holding capacity was determined as in a previous study⁶⁸. Soil nitrogen was determined using a CN analyser. Soil phosphorus, potassium and magnesium concentrations were determined using inductively coupled plasma (ICP) spectroscopy after acid digestion⁶⁹. Total soil organic carbon content was determined using a CN analyser (after removing soil carbonates) and wet-chemistry methods⁵⁰. The proportion of soil-borne fungal potential plant pathogens and fungal plant-soil mutualistic organisms (arbuscular and ectomycorrhizal fungi) were determined as the sum of all taxa classified as such from FUNGuild⁷⁰ using our 18S dataset. We found FUNGuild information for 297 ASVs of arbuscular mycorrhizal fungi, 217 ASVs of ectomycorrhizal fungi and 165 ASVs of soil-borne potential fungal plant pathogens. Pest control was calculated as the inverse of the proportion of soil-borne potential plant pathogens ($-1 \times$ proportion) as described before⁶⁷. Thus, locations with higher levels of pest control also have lower proportions of plant pathogens. We only focused on those taxa that support unique trophic lifestyles. The activity of phosphatase (phosphorus mineralization), β -glucosidase (starch degradation) and N -acetyl- β -glucosaminidase (chitin degradation) was determined as before⁷¹, using a high-throughput fluorescence microplate method. The exact amount of available information might differ for different ecosystem services (see Supplementary Table 2). The total number of samples available for each soil attribute is provided in Supplementary Table 2. We calculated ecosystem services as the standardized (0–1) average of soil attributes within each ecosystem service (for example, fertility: nitrogen, phosphorus, potassium and magnesium; mutualism: arbuscular and ectomycorrhizal fungi; organic matter decomposition: phosphorus mineralization, chitin and starch degradation) using a multifunctionality approach⁷². Furthermore, we acknowledge that the number and type of ecosystem services considered here might be limited to characterize the range of ecological functions driven by soil communities. Therefore, for a subset of the data for which other variables are available, we correlated our ecosystem services to additional information on carbon content, enzymes, nutrient availability from IEMS (a proxy of nitrogen mineralization⁷³) and metagenomics (see Supplementary Table 10).

Environmental data

Elevation and climatic information for each location was obtained from WorldClim v.2 (1-km² resolution; <https://www.worldclim.org/data/bioclim.html>), including information on climatologies and on the seasonality of temperature and precipitation. Soil pH was determined with a soil pH meter from a soil and water mix⁴⁹. Texture was determined as in a previous report⁴⁹ and, in the case of missing information, this was filled using Soilgrid v.2 (<https://soilgrids.org>; as in ref. ¹⁶).

<p>Information on dominant vegetation (forest, shrublands or grasslands) was obtained as part of the field survey.</p> <h3>Drivers of soil biodiversity and services</h3> <p>To investigate the environmental factors associated with soil biodiversity and services, we first used machine learning random forest modeling⁷⁴. We used the R package 'rfpermute' to conduct these analyses. To further strengthen our results, we repeated the same analysis using XGboost algorithms^{75,76}, which allow for fine-tuning of the model outcomes. In brief, we used a gbtree booster using the root-mean-square error (RMSE) as an objective function using <i>k</i>-fold cross-validation. We tuned the learning rate (the ETA parameter in XGboost), min_chld_weight, max depth, resample, gamma and nfold⁷⁷ using a Bayesian optimization approach with the package 'ParBayesianOptimization' in R⁷⁸. This allows for efficient implementation of an optimizing search. The parameter that we decided to optimize was the test RMSE (rather than the train one), to prevent overfitting⁷⁹. The number of trees fitted was also estimated on this criterion using the function xgb.cv from the package xgboost in R with 10% of data as the validation set. As feature importance, we extracted the gain obtained in predictive, using the function xgb.importance (the results for this second analysis can be found in Supplementary Fig. 7 and Supplementary Tables 11 and 12). Here we chose to train the resulting models using the function xgcv, which fits from 1 to 200 trees to the data and compares trained models to the cross-validation set. We chose to use the number of trees that minimized the test RMSE to prevent overfitting. In the new analyses, we parameterized eight key hyperparameters, including the number of trees (nroud) and the number of features (colsample_bytree). A full table of the resulting parameters is given in Supplementary Table 11. To further strength our analysis, we compared the results that were obtained with our approach with results that were obtained using a GAMM model (considering random factors). GAMMs were performed by flooring coordinates (latitude and longitude) and using their combination as a random factor: ("mdl=gamm(data = ddi, formula = y~s(Latitude)+s(Longitude.cosine)+s(Longitude.sine)+s(s_elev)+Forest+Grassland+Shrubland+s(s_MAT)+s(s_TSEA)+s(s_PSEA)+s(s_MAP)+s(s_SOC)+s(s_Texture)+s(s_pH) random = list(RF = -1))"; RF is the floored combination of latitude and longitude: "dd\$RF=as.factor(paste(floor(coords\$long),floor(coords\$lat)))"). This approach allows to control for spatial autocorrelation, apart from the nested structure of our data. Our results show (Supplementary Fig. 10) that the GAMM model provided highly correlated results (Spearman correlations of greater than 0.7) for all variables considered and almost identical (very aligned to 1:1 line) predictions to the machine learning algorithms used in our manuscript, suggesting that the type of modelling (for example, random forest versus GAMM) does not influence our results and conclusions. We also performed Spearman correlations to better describe the direction of the relationship between environmental factors and soil biodiversity and services. We also correlated all soil biodiversity and services attributes looking for potential trade-offs using Spearman rank correlations. All the analyses in this section are non-parametric, and are especially recommended when dealing with both linear and non-linear relationships. Analyses were done at the sample level to account for within-location variation in soil properties, biodiversity and services.</p> <h3>Global hotspots of soil biodiversity and services</h3> <p>We used spatially explicit random forest models to predict the distribution of each soil biodiversity and ecosystem service variable. We were able to do these spatial analyses for three main reasons: (1) the high-quality standardized biodiversity and ecosystem service dataset in which biodiversity and services are measured for the same samples, and analysed using the same protocols; (2) biodiversity and ecosystem services were highly correlated with key environmental factors at the global scale (Extended Data Figs. 2 and 3); and (3) the large gradient of environmental conditions in our global dataset covers a large portion (61.6% based on a Mahalanobis analysis) of the large-scale environmental variability of the planet. Note that we further spatially constrained our analyses to exclude all environmental outliers^{22,80}.</p>	<p>To map each soil biodiversity and ecosystem service variable we used spatially explicit random forest models. For that, we used ArcGIS Pro, which estimates random forest models by using an adaptation of the random forest algorithm (a supervised machine learning regression approach) proposed previously^{74,81,82}. Forest-based regressions were trained on the basis of 90% of the dataset; the remaining 10% of the dataset was used for validation purposes. Regression training and validation parameters are given in Supplementary Table 5 (for alpha diversity and community dissimilarity) and Supplementary Table 6 (for ecosystem services). The fitted prediction model was then used to predict the unknown space using a prediction dataset that included all environmental explanatory factors; that is, elevation, carbon, pH, fine texture, mean annual temperature and precipitation, temperature and precipitation seasonality, forest, grassland and shrubland. In the case of the analyses related to the ecosystem services, carbon was excluded from all the models. All models were fitted using 1,000 runs for validation and fitting. Before prediction all variables included in the dataset and the predictors were resampled to 0.25 degrees using an average estimator and scaled. All predictions were made using a 0.25×0.25 degree pixel size. All environmental variables that were used for spatial projection are listed in Supplementary Table 14.</p> <p>Global hotspots were then calculated using a Getis-Ord Gi* spatial clustering method^{83–85}. The Getis-Ord Gi* statistic was calculated for each location (0.25×0.25 degree pixel) in the dataset. The resulting z-scores were used to estimate whether a given location has statistically high or low values and whether these values are spatially clustered. This is done by assessing each location within the context of neighbouring locations. Statistically significant positive z-scores indicate clustering of high values (hotspot) and statistically significant negative z-scores indicate the clustering of low values (coldspot). Values for classifying hotspots (positive z-scores) for each variable were taken from the 99% confidence interval. Getis-Ord allows the use of the false discovery rate (FDR) correction, which was also applied here, and adjusts the statistical significance of a hotspot detection to account for multiple testing (with a confidence level of 0.95) and spatial dependency⁸⁶. This analysis resulted in a hotspot map for each combination; that is, five hotspot maps for alpha diversity, five for community dissimilarity and six for ecosystem services (Extended Data Figs. 4 and 5). We then overlaid the maps for each ecological dimension (by summing the hotspot maps for each variable in each ecological dimension: alpha diversity, community dissimilarity and ecosystem services) to obtain a global representation of soil biodiversity hotspots, in which a high value corresponds to a concentration of hotspots across multiple taxa or ecosystem services (Fig. 1a–c).</p> <h3>Spatial uncertainty estimations</h3> <p>One of the difficulties of performing prediction of response variables using a new input dataset is the fact that the new input environmental values might differ substantially from values used to estimate the models. Therefore, estimating uncertainties on the environmental coverage of the datasets as well as the estimations of both biodiversity and ecosystem services is a complex but necessary requirement in such scenario modelling approaches⁸⁷. For this, we implemented a two-stage approach to tackle both the assessment of the environmental representation of the soil biodiversity and ecosystem services dataset used and the uncertainty related to the estimation of each variable or group of variables. Regarding the first, we calculated the Mahalanobis distance in multidimensional space (here considering the 12 dimensions given by the environmental variables used for modelling (that is, elevation, carbon, nitrogen, pH, fine texture, mean annual temperature and precipitation, temperature and precipitation seasonality, forest, grassland and shrubland) and centred on the known distribution given</p>
---	---

Article

by the characteristics, for the same environmental variables, of the soil biodiversity and ecosystem service dataset. This analysis calculates the distance of any point in space to the statistical centre, given by the multivariate mean (considering all environmental variables used) of the known distribution. It is often used to detect outliers in point cloud distributions that are assumed to follow a multivariate normal distribution^{80,88}. The Mahalanobis distance follows a chi-squared distribution with d degrees of freedom, where d is the dimension of the multidimensional space ($d = 12$ in our case). Environmental outliers were estimated for a chi-squared value of 0.9 (areas in grey in Supplementary Fig. 5).

Although this distance is an informative measure of how close a new data point is to the distribution of points in space used to estimate each model, we used a second measure to assess the spatial uncertainty of the estimated values for each model. For this analysis, for each soil biodiversity and ecosystem service variable, we calculated 1,000 random iterations of each random forest model and estimated the upper and lower 25% quantile of the distribution of values. We then evaluated uncertainty as the difference between the upper and the lower level of the iteration space for each individual variable. An average representation for each dimension is given in Supplementary Fig. 5.

Projections of soil biodiversity and ecosystem services under global change scenarios

For the projections of soil biodiversity and ecosystem services, we used the available datasets from the Inter-Sectoral Impact Model Inter-comparison Project (ISIMIP)⁸⁹ and from the Land-Use Model Inter-comparison Project (LUMIP)⁹⁰, both of which are projects from the Intergovernmental Panel for Climate Change (IPCC). The selection of scenarios followed the protocol described in a previous study⁹¹.

In terms of climate change projections, we used a bias-corrected future projections dataset for both precipitation and temperature related variables⁸⁹. We considered three representative concentration pathways RCP2.6, RCP6.0 and RCP8.5 (ref. ⁸⁹) with forcing data from three different general circulation models, the IPSL-cm5a-lr, gfdl-esm2m and noresm1-m (ref. ⁹²). For land-use projections, we used the dataset provided by the land-use Harmonized v.2.0 project (<http://luh.umd.edu/>) (refs. ^{28,93,94}). This dataset was produced in the context of the World Climate Research Program Coupled Model Intercomparison Project 6 (CMIP6)^{28,45,95,96} and contains a harmonized set of land-use scenarios that are consistent between historical reconstructions and future projections. These modelled projections reproduce annual land-use reconstructions for different integrated assessment models (IAMs) and shared socio-economic pathways (SSPs, from 2015 to 2100) at 0.25-degree resolution, which were developed and widely used to support future biodiversity projections^{97,98}. These shared socio-economic pathways represent a range of plausible futures based on different socio-economic challenges for climate change mitigation (low in SSP1 and SSP4; high in SSP3 and SSP5), and potential challenges for adaptation (low in SSP1 and SSP5; high in SSP3 and SSP4). Full descriptions of these pathways and scenarios have been published previously²⁸, and we provide here a summary of the main characteristics (based on a previous report⁴⁵).

SSP1. In SSP1, the world shifts gradually, but pervasively, toward a more sustainable path, with its focus on achieving the global development goals, increasing environmental awareness and a gradual move toward less-resource-intensive societies. At present, emerging economies have followed the resource-intensive development model of industrialized countries, but SSP1, with the focus on equity, and the de-emphasis of economic growth as a goal in high-income countries, leads industrialized countries to support developing countries in their development goals, including green growth strategies, by providing access to human and financial resources and new technologies.

SSP3. In SSP3, a resurgent nationalism, concerns about competitiveness and security, and regional conflicts push countries to increasingly

focus on domestic or, at most, regional issues. International fragmentation and a world characterized by regional rivalry can already be seen in some of the current regional rivalries and conflicts, but contrasts with globalization trends in other areas. Regional conflict over territorial or national issues produces larger conflicts between major countries, giving rise to increasing antagonism between and within regional blocs, reducing support for international institutions and weakening progress toward the global development goals, particularly in some middle-income countries.

SSP4. In SSP4, highly unequal investments in human capital, combined with increasing disparities in economic opportunity and political power, lead to increasing inequalities and stratification both across and within countries. Both across- and within-country inequality is assumed to arise from biased technology development, generally low and highly unequal investments in education resulting in increased restricted access, and reinforced wealth inequality. This pathway assumes that growth is substantially smaller than it is today, but does not assume that it is halted entirely. It also assumes an increased conflict over energy resources between consuming countries and producing countries, particularly if resources are further constrained.

SSP5. In SSP5, there is a foreseen acceleration in globalization and rapid development of developing countries. The digital revolution enables an enhanced global discourse, which may lead to a rapid rise in global institutions and promote the ability for global coordination. This pathway is driven by the economic success of industrialized and emerging economies to produce rapid technological progress and the development of human capital as the path to sustainable development. Global markets are increasingly integrated, with the push for economic and social development coupled with the exploitation of abundant fossil fuel resources and the adoption of resource and energy-intensive lifestyles around the world. All of these factors lead to rapid growth of the global economy. There is the ability to effectively manage social and ecological systems, including by geo-engineering if necessary. Although local environmental effects are addressed effectively by technological solutions, there is relatively little effort to avoid potential global environmental effects owing to a perceived trade-off with progress on economic development.

All temporal changes (2070 minus 2015, using forecasting predictions) were calculated using 2015 as a baseline to which all future predictions were compared.

Reporting summary

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

Data availability

All of the materials, raw data and protocols used in this Article are available upon request and without restriction, and all data are publicly available at <https://doi.org/10.6084/m9.figshare.20221713>.

49. Maestre, F. T. et al. Plant species richness and ecosystem multifunctionality in global drylands. *Science* **335**, 214–218 (2012).
50. Delgado-Baquerizo, M. et al. Changes in belowground biodiversity during ecosystem development. *Proc. Natl Acad. Sci. USA* **116**, 6891–6896 (2019).
51. Mace, G. M. Whose conservation? *Science* **345**, 1558–1560 (2014).
52. Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. & Huse, S. M. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One* **4**, e6372 (2009).
53. Stoeck, T. et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
54. Ramirez, K. S. et al. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. Biol. Sci.* **281**, 20141988 (2014).
55. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
56. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

57. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31**, 3476–3482 (2015).
58. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. Preprint at *bioRxiv* <https://doi.org/10.1101/081257> (2016).
59. Tedersoo, L. et al. Towards understanding diversity, endemism and global change vulnerability of soil fungi. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.03.17.484796> (2022).
60. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
61. Delgado-Baquerizo, M. et al. Global homogenization of the structure and function in the soil microbiome of urban greenspaces. *Sci. Adv.* **7**, eabg5809 (2021).
62. Phillips, H. R. P., Heintz-Buschart, A. & Eisenhauer, N. Putting soil invertebrate diversity on the map. *Mol. Ecol.* **29**, 655–657 (2020).
63. Xiong, W. et al. A global overview of the trophic structure within microbiomes across ecosystems. *Environ. Int.* **151**, 106438 (2021).
64. Drummond, A. J. et al. Evaluating a multigene environmental DNA approach for biodiversity assessment. *Gigascience* **4**, 46 (2015).
65. Oliverio, A. M., Gan, H., Wickings, K. & Fierer, N. A DNA metabarcoding approach to characterize soil arthropod communities. *Soil Biol. Biochem.* **125**, 37–43 (2018).
66. Horton, D. J., Kershner, M. W. & Blackwood, C. B. Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes. *Eur. J. Soil Biol.* **80**, 43–48 (2017).
67. Delgado-Baquerizo, M. et al. Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nat. Ecol. Evol.* **4**, 210–220 (2020).
68. Carter, M. R. & Gregorich, E. G. (eds) *Soil Sampling and Methods of Analysis* (CRC Press, 2007).
69. Sparks, D. L. et al. (eds) *Methods of Soil Analysis, Part 3: Chemical Methods* (Wiley, 2020).
70. Nguyen, N. H. et al. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).
71. Bell, C. W. et al. High-throughput fluorometric measurement of potential soil extracellular enzyme activities. *J. Vis. Exp.* **81**, e50961 (2013).
72. Wang, L. et al. Diversifying livestock promotes multidiversity and multifunctionality in managed grasslands. *Proc. Natl Acad. Sci. USA* **116**, 6187–6192 (2019).
73. Durán, J., Delgado-Baquerizo, M., Rodríguez, A., Covelo, F. & Gallardo, A. Ionic exchange membranes (IEMs): a good indicator of soil inorganic N production. *Soil Biol. Biochem.* **57**, 964–968 (2013).
74. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
75. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
76. Sharma, N. *XGBoost: The Extreme Gradient Boosting for Mining Applications* (GRIN Verlag, 2018).
77. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, 2016).
78. Wilson. ParBayesianOptimization: Parallel Bayesian Optimization of Hyperparameters. R version 1 <https://CRAN.R-project.org/package=ParBayesianOptimization> (2021).
79. Hastie, T., Friedman, J. & Tibshirani, R. *The Elements of Statistical Learning* (Springer, 2001).
80. Jackson, D. A. & Chen, Y. Robust principal component analysis and outlier detection with ecological data. *Environmetrics* **15**, 129–139 (2004).
81. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
82. Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. *Classification and Regression Trees* (Routledge, 1984).
83. Ord, J. K. & Getis, A. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr. Anal.* **27**, 286–306 (2010).
84. Getis, A. & Ord, J. K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **24**, 189–206 (2010).
85. Prasannakumar, V., Vijith, H., Charutha, R. & Geetha, N. Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia Soc. Behav. Sci.* **21**, 317–325 (2011).
86. Lin, G. Comparing spatial clustering tests based on rare to common spatial events. *Comput. Environ. Urban Syst.* **28**, 691–699 (2004).
87. Araújo, M. B. et al. Standards for distribution models in biodiversity assessments. *Sci. Adv.* **5**, eaat4858 (2019).
88. Rousseeuw, P. J. & van Zomeren, B. C. Unmasking multivariate outliers and leverage points. *J. Am. Stat. Assoc.* **85**, 633–639 (1990).
89. Hempel, S., Frieler, K., Warszawski, L., Schewe, J. & Piontek, F. A trend-preserving bias correction—the ISI-MIP approach. *Earth Syst. Dyn.* **4**, 219–236 (2013).
90. Lawrence, D. M. et al. The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design. *Geosci. Model Dev.* **9**, 2973–2998 (2016).
91. Kim, H. et al. A protocol for an intercomparison of biodiversity and ecosystem services models using harmonized land-use and climate scenarios. *Geosci. Model Dev.* **11**, 4537–4562 (2018).
92. Dufresne, J.-L. et al. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP6. *Clim. Dyn.* **40**, 2123–2165 (2013).
93. Hurtt, G. C. et al. Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Clim. Change* **109**, 117 (2011).
94. Hurtt, G. C. et al. Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6. *Geosci. Model Dev.* **13**, 5425–5464 (2020).
95. Riahi, K. et al. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Glob. Environ. Change* **42**, 153–168 (2017).
96. O'Neill, B. C. et al. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Clim. Change* **122**, 387–400 (2014).
97. Newbold, T. et al. Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015).
98. Powers, R. P. & Jetz, W. Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. *Nat. Clim. Chang.* **9**, 323–329 (2019).

Acknowledgements We thank all of the researchers who were involved in the collection of field data. This project received funding from the British Ecological Society (agreement LRA171193; MUSGONET). C.A.G. and N.E. were funded by DFG-FZT 118, 202548816; C.A.G. was supported by FCT-PTDC/BIA-CBI/2340/2020; M.D.-B. was supported by RYC2018-025483-I, PID2020-115813RA-I00/MCIN/AEI/10.13039/50100011033 and P20_00879. M.A.-M. and S.A. were funded by FONDECYT 1181034 and ANID-PIA-Anillo INACH ACT192057. J.D. and A.R. acknowledge support from IF/00950/2014, 2020,03670.CEECIND, SFRH/BDP/108913/2015 and UIDB/04004/2020. Y.-R.L. was supported by 2662019PY010 from the FRCU. L.T. was supported by the ESF grant PRG632. F.B. and J.L.M. were supported by i-LINK+2018 (LINKA20069) funded by CSIC. C.T.-D. was supported by the Grupo de Biodiversidad y Cambio Global UBB-GI 170509/EF. C.P. was supported by the EU H2020 grant agreement 101000224. H.C. was supported by NSFC32101335, FRFCU2412021QD014 and CPSF2021M690589. J.P.V. was supported by DST (DST/INT/SL/P-31/2021) SERB (EEQ/2021/001083) and BHU-IoE (6031).

Author contributions C.A.G. and M.D.-B. developed the original idea of the analyses presented in the manuscript. M.D.-B. designed the field study and wrote the grant that funded the work. Field data were collected by M.B., S.A., F.D.A., A.R.B., J.L.B.-P., A.d.l.R., J.D., T.G., J.G.I., Y.-R.L., T.P.M., S.M., M.A.M.-M., A.M., T.U.N., G.F.P.-B., C.P., J.P.V., A. Rey, A. Rodríguez, A.L.T., C.T.-D., P.T., L.W., Jianyong Wang, E.Z., X.Z., X.-Q. Z. and M.D.-B. Laboratory analyses were done by M.D.-B., H.C., F.B., J.L.M., S.P. and L.T. Statistical analyses, mapping and ecological modelling were done by C.A.G., M.D.-B. and M.B. Bioinformatic analyses were done by B.K.S. and Juntao Wang. The manuscript was written by C.A.G. and M.D.-B. and edited by N.E. and D.J.E., with contributions from all co-authors.

Competing interests The authors declare no competing interests.

Additional information

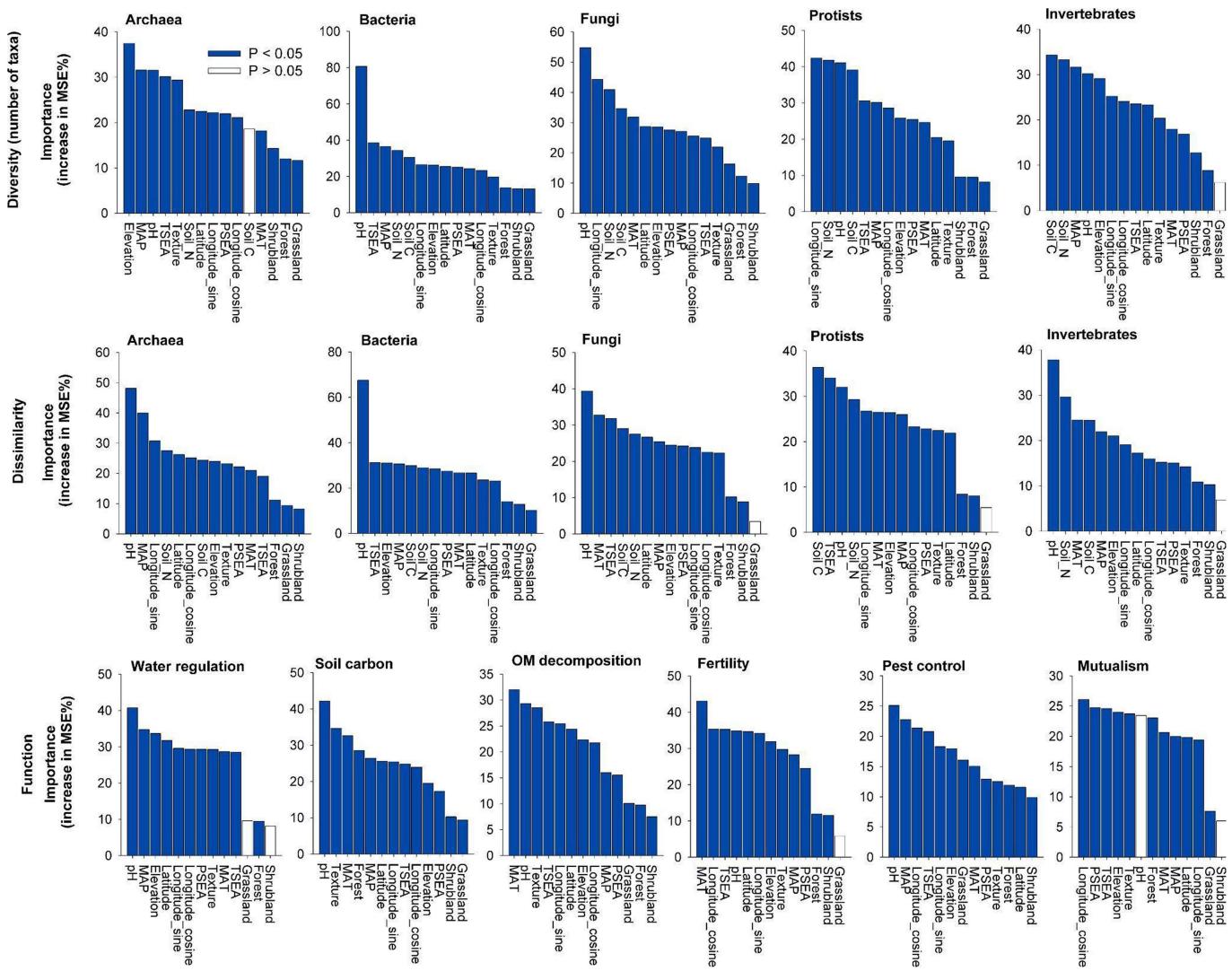
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-05292-x>.

Correspondence and requests for materials should be addressed to Carlos A. Guerra or Manuel Delgado-Baquerizo.

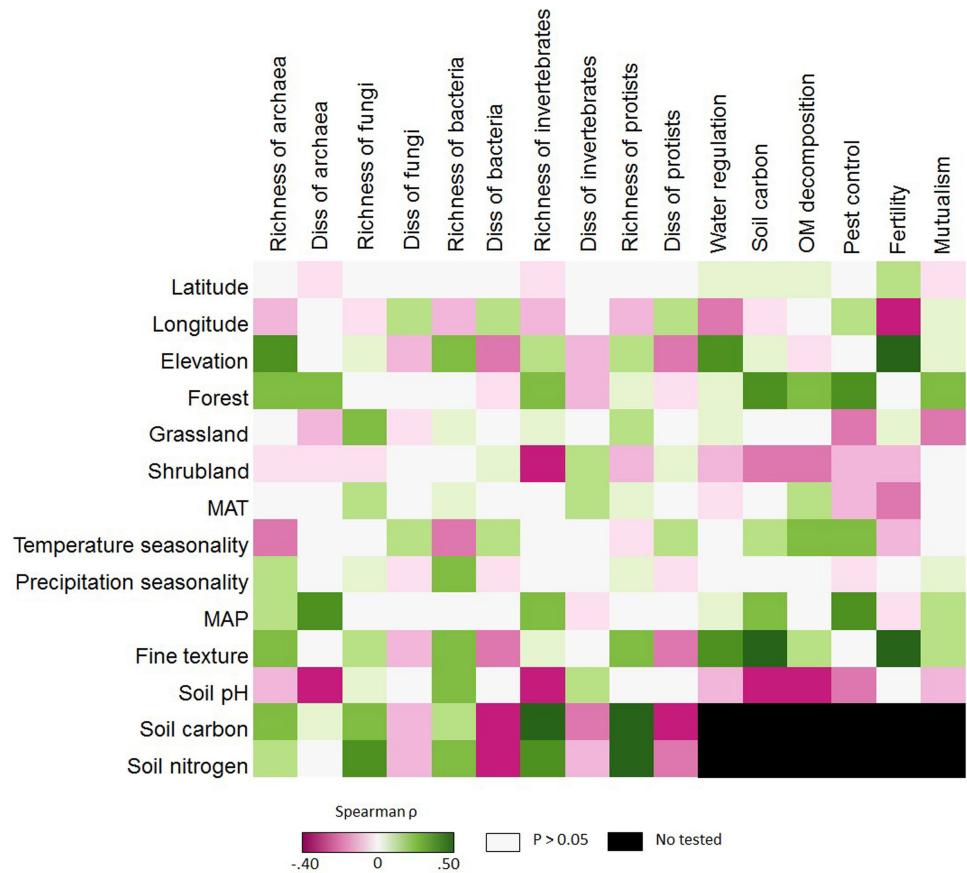
Peer review information *Nature* thanks Peter de Ruiter, Ruhollah Taghizadeh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

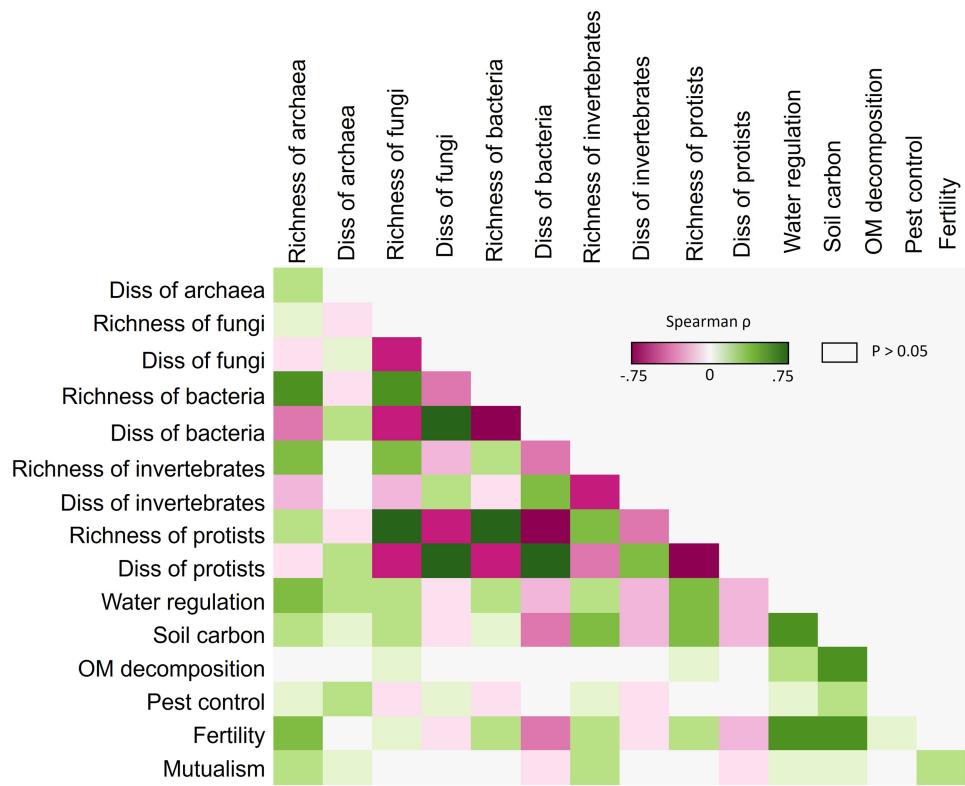
Article



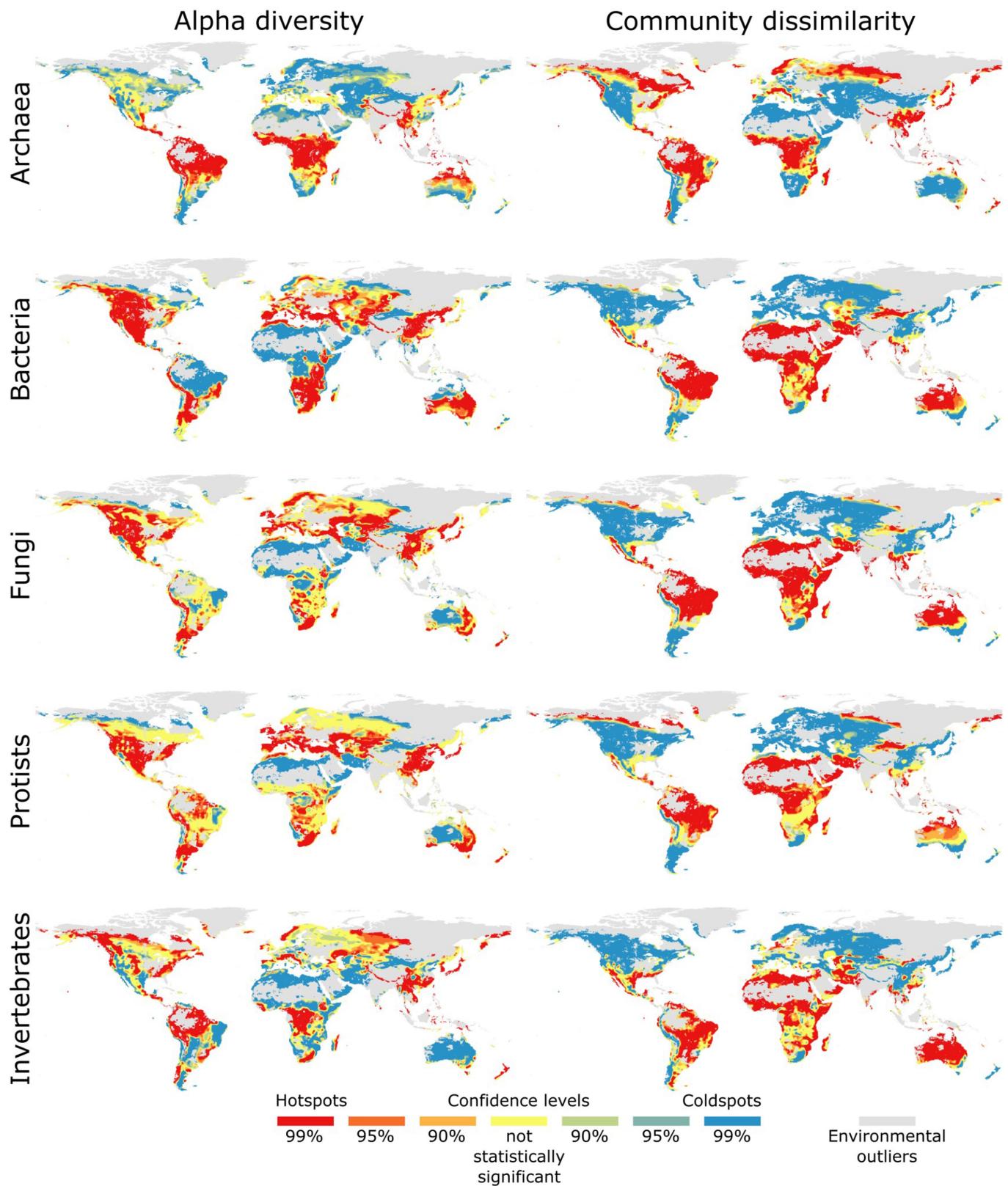
Extended Data Fig. 1 | Results of the random forest analysis to identify the main environmental factors associated with soil biodiversity and ecosystem services. Random forest analyses were done using the rfPermute function of the R package with the same name. MSE, mean square error.



Extended Data Fig. 2 | Spearman correlations between environmental factors and soil biodiversity and ecosystem services. *N* in Supplementary Table 1.



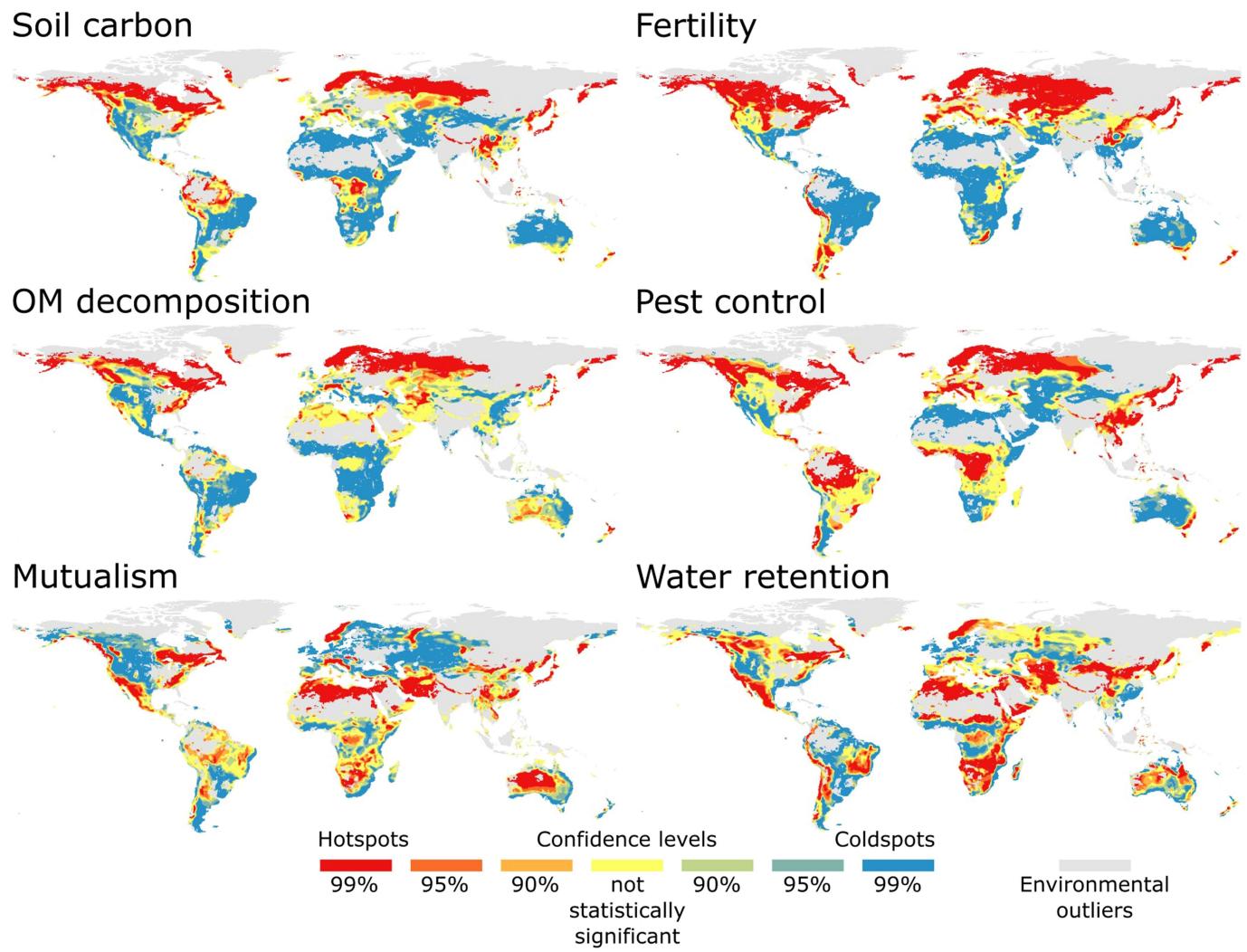
Extended Data Fig. 3 | Spearman correlations between soil biodiversity and ecosystem services. Total n-values in Supplementary Table 1.



Extended Data Fig. 4 | Hotspot and coldspot maps for alpha diversity (left) and community dissimilarity (right). The Getis-Ord G_i^* statistic was calculated for each location (0.25×0.25 deg pixel size) in the dataset 1–3. The resulting z-scores were used to estimate if a given location has statistically high or low values and if these values are spatially clustered. This is done by

assessing each location within the context of neighbouring locations. Statistically significant positive z-scores indicate clustering of high values (hotspot) and statistically significant negative z-scores the clustering of low values (coldspot). Values are plotted for both positive (hotspots) and negative (coldspots) 99%, 95%, and 90% confidence levels.

Article



Extended Data Fig. 5 | Hotspot and coldspot maps for ecosystem services: soil carbon, fertility, organic matter decomposition, pest control, mutualism and water retention. The Getis-Ord Gi* statistic was calculated for each location (0.25×0.25 deg pixel size) in the dataset 1–3. The resulting z-scores were used to estimate if a given location has statistically high or low values and if these values are spatially clustered. This is done by assessing each

location within the context of neighbouring locations. Statistically significant positive z-scores indicate clustering of high values (hotspot) and statistically significant negative z-scores the clustering of low values (coldspot). Values are plotted for both positive (hotspots) and negative (coldspots) 99%, 95%, and 90% confidence levels.

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

We used composite topsoil samples from global field surveys which were conducted between 2016–2019 following standardized field protocols. This global field survey includes 151 locations from all continents and 23 countries, from which 615 composite topsoil samples were selected, providing a large representation of all climatic and vegetation biomes in the planet (Supplementary Fig. 1). Between three and five composite soil (top ~10 cm) samples (from 5–10 soil cores) were collected in these locations. Environmental predictors were collected from publicly available datasets. Soil properties, biodiversity and ecosystem services were measured as explained in the Method section of our manuscript.

Data analysis

The richness (number of phylotypes) and dissimilarity (averaged Jaccard distance across samples from presence/absence matrices to account for dissimilarity in species, rather than in their proportions) of archaea, bacteria, fungi, protists and invertebrates was determined using amplicon sequencing technology (Illumina MiSeq platform). Soil DNA was extracted using the Powersoil® DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. We conducted additional analyses to provide evidence that our choice of rarefaction level did not affect our results or conclusions. Here, using the samples with the highest sequence/sample yield, we tested for the impact of different levels of rarefaction on soil biodiversity. We provide additional evidence that primer sets are not influencing the global patterns reported here. For a subset of samples, we generated additional molecular information for fungal (ITS PacBio sequencing; ITS9mun/ITS4ngsUni primer sets) and bacterial (16s rRNA Miseq Sequencing; 341F/805R primer sets) data. Finally, to provide further evidence that 18s rRNA Miseq Sequencing can in this case provide a solid representation of the global patterns in soil-borne mycorrhizal fungi and fungal potential plant pathogens, we compare the global patterns (see mapping method below) in the proportion of soil-borne mycorrhizal fungi and fungal potential plant pathogens determined using 18s rRNA Miseq Sequencing with the subset of data including ITS PacBio sequencing. To investigate the environmental factors associated with soil biodiversity and services, we first used machine learning Random Forest modeling. We used the R package "rfpermute" to conduct these analyses. We also conduct Spearman correlations to better describe the direction of the relationship between environmental factors and soil biodiversity and services. Next, we used spatially explicit random forest models to predict the distribution of each soil biodiversity and ecosystem service variable. To map each soil biodiversity and ecosystem service variable we used spatially explicit random forest models. For that, we used ArcGIS Pro that estimates random forest models by using an adaptation of the random forest algorithm (a supervised machine learning regression approach) proposed by Breiman et al..

Forest-based regressions were trained based on 90% of the dataset, the remaining 10% of the dataset were used for validation purposes. All models were fitted using 200 decision trees and 1000 runs for validation and fitting. Prior to prediction all variables included in the dataset and the predictors were scaled and predictions were made using a 0.25x0.25 deg. pixel size. Global hotspots were then calculated using a Getis-Ord Gi* spatial clustering method. The Getis-Ord Gi* statistic was calculated for each location (0.25x0.25 deg. pixel) in the dataset. The resulting z-scores were used to estimate if a given location has statistically high or low values and if these values are spatially clustered. Finally, we have implemented a two stage approach to tackle both the assessment of the environmental representation of the soil biodiversity and ecosystem services dataset used and the uncertainty related to the estimation of each variable or group of variables. For this, we calculated the mahalanobis distance in multidimensional space to assess environmental coverage and a second measure to assess the spatial uncertainty of the estimated values for each model. In order to do this analysis, for each soil biodiversity and ecosystem service variable, we calculated 1000 random iterations of each random forest model and estimated the upper and lower 25% quantile of the distribution of values.

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 the materials, raw data, and protocols used in the article are available upon request and without restriction, and all data will be made publicly available in a public repository: <https://doi.org/10.6084/m9.figshare.20221713.v3>

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

This study is based in a global field survey includes 151 locations from all continents and 23 countries, from which 615 composite topsoil samples, providing a large representation of all climatic and vegetation biomes in the planet. Based on this survey, we aim to identify the global hotspots of soil biodiversity conservation as well as estimating the potential threats from climate and land use change.

Research sample

In each sampling site a composite (from 5-10 soil cores) top soil sample (top ~10 cm) was collected in an homogeneous location in order to represent the diversity within the site in an adequate way. Overall we used 615 composite topsoil samples including a wide characterization of their soil biodiversity and ecosystem functions related to important ecosystem services.

Sampling strategy

We used composite topsoil samples from global field surveys which were conducted between 2016-2019 following standardized field protocols. This global field survey includes 151 locations from all continents and 23 countries, from which 615 composite topsoil samples were selected, providing a large representation of all climatic and vegetation biomes in the planet (Supplementary Fig. 1). Between three and five composite soil (top ~10 cm) samples (from 5-10 soil cores) were collected in these locations following the protocol described in Maestre et al. (2012).

Data collection

The richness (number of phylotypes) and dissimilarity (averaged Jaccard distance across samples from presence/absence matrices to account for dissimilarity in species, rather than in their proportions) of archaea, bacteria, fungi, protists and invertebrates was determined using amplicon sequencing technology (Illumina MiSeq platform) following the protocol in Delgado-Baquerizo et al. (2019). Soil DNA was extracted using the Powersoil® DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. Environmental predictors were collected from publicly available datasets. Elevation and climatic information for each location was obtained from WorldClim v2 (<https://www.worldclim.org/data/bioclim.html>), including information on climatologies and on the seasonality of temperature and precipitation. Soil pH was determined with a soil pH-meter from a soil-water mix. Texture was determined as in Maestre et al. and, in the case of missing information, this was filled using Soilgrid v2 (<https://soilgrids.org>). Information on dominant vegetation (forest, shrublands or grasslands) was obtained as part of the field survey. For the projections of soil biodiversity and ecosystem services, we used the available datasets from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) and from the land-use Model Intercomparison Project (LUMIP) both activities from the Intergovernmental Panel for Climate Change (IPCC). The selection of scenarios followed the protocol laid out by Kim et al. (2018).

Timing and spatial scale

Field surveys which were conducted between 2016-2019.

Data exclusions

There were no data exclusions.

Reproducibility

Within the manuscript, we clearly state all the steps taken to ensure the reproducibility of the study including describing and following standard sampling and analytical protocols and identifying all code packages and languages used.

Randomization

For the estimation of the spatial uncertainties related to our projections we calculated 1000 random iterations of each random forest model and estimated the upper and lower 25% quantile of the distribution of values. Similarly, in each model run we used 10% of the data (selected randomly) for validation.

Blinding

Blinding was not relevant for this study

Did the study involve field work? Yes No

Field work, collection and transport

Field conditions

Samples were all collected in terrestrial systems.

Location

Global

Access & import/export

Samples were collected by authors in their respective locations and using local permits.

Disturbance

No significant disturbance was inflicted to the sampled sites given the type of sampling method implemented.

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 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |