Enumerating soil biodiversity

Mark A. Anthony[a,b,1] (ID), S. Franz Bender[a,c] (ID), and Marcel G. A. van der Heijden[a,c,1]

Soil is an immense habitat for diverse organisms across the tree of life, but just how many organisms live in soil is surprisingly unknown. Previous efforts to enumerate soil biodiversity consider only certain types of organisms (e.g., animals) or report values for diverse groups without partitioning species that live in soil versus other habitats. Here, we reviewed the biodiversity literature to show that soil is likely home to 59 ± 15% of the species on Earth. We therefore estimate an approximately two times greater soil biodiversity than previous estimates, and we include representatives from the simplest (microbial) to most complex (mammals) organisms. Enchytraeidae have the greatest percentage of species in soil (98.6%), followed by fungi (90%), Plantae (85.5%), and Isoptera (84.2%). Our results demonstrate that soil is the most biodiverse singular habitat. By using this estimate of soil biodiversity, we can more accurately and quantitatively advocate for soil organismal conservation and restoration as a central goal of the Anthropocene.

biodiversity | diversity | soil biodiversity | soil diversity

"How many species on Earth live in soil?" Like many simple questions, it is one of the most challenging to answer. A timely account of soil biodiversity is critical as Earth faces another wave of mass extinctions (1), and evidence points to humans as the major cause (2–4). Curtailing and reversing this trend requires biodiversity monitoring and conservation programs like the Endangered Species Act in the United States and the Global IUCN Redlist, the most extensive source of information on global extinction risks. Yet, these efforts largely exclude species inhabiting soil, despite the critical importance of soil organisms to nearly every Earth function (5–7). Only with a complete estimation of the quantity of life that lives in soil can we understand the magnitude of value for conserving and restoring soil biodiversity.

Soil organisms are indispensable drivers of ecosystem composition and function. They govern global biogeochemical fluxes and directly influence rates of climate change and human health (8–10). Without an accurate estimate of soil biodiversity, we not only overlook a fundamental component of global biodiversity but additionally lack the quantitative information essential for policy advocacy (11). There has been one previous estimate of global soil biodiversity based on soil animals. Decaëns et al. (12) reported that at least 25% of described animal species live in or on soil based on a rapid survey of known animal species from a limited set of encyclopedic sources (12). This value has been widely taken up in soil biodiversity research, but it has also been incorrectly cited to comprehensively represent all species versus just animals. To systematically estimate soil biodiversity, we must synthesize species numbers across the tree of life (see organisms living in soil in Fig. 1) and overcome myriad theoretical and technical obstacles.

The first main obstacle to synthesizing global biodiversity is understanding the "species" concept across different life forms. Most familiar are the biological and evolutionary species concepts which define species as populations incapable of effectively mating with other groups (13) or as separately evolving lineages with different heritable traits at the population level (14). There are secondary species concepts that better account for all types of organisms using operational methods to distinguish species, including asexually and nonreproducing groups (15). Differences among these concepts are most salient when comparing macrobes to microbes. For example, whereas mammal species diversity is often based on direct observations in the field paired with information on species distributions, traits, and molecular analyses (16), microbial species diversity has been quantified primarily using molecular analyses of DNA hybridization (17), ribosomal RNA gene polymorphism (18), and/or dissimilarity in other regions of the microbial genome (19). Microbial species differences are also blurred by horizontal gene transfer (20) and/or asexual reproduction (21), but operationally, the vast majority of researchers define microbial species based on sequence similarity of ribosomal RNA regions (22–24), enabling a fairly straightforward integration across microbial studies (25). The species

Significance

Soil organisms mediate unique functions we rely on for food, fiber, and human and planetary health. Despite the significance of soil life, we lack a quantitative estimate of soil biodiversity, making it challenging to advocate for the importance of protecting, preserving, and restoring soil life. Here, we show that soil is likely home to 59% of life including everything from microbes to mammals, making it the singular most biodiverse habitat on Earth. Our enumeration can enable stakeholders to more quantitatively advocate for soils in the face of the biodiversity crisis.

Author affiliations: [a]Plant-Soil Interactions Unit, Research Division Agroecology and Environment, Agroscope, Zürich 8046, Switzerland; [b]Swiss Federal Institute for Forest, Snow and Landscape Research, Forest Dynamics Research Unit, Birmensdorf 8903, Switzerland; and [c]Department of Plant and Microbial Biology, University of Zürich, Zürich 8008, Switzerland

[1]To whom correspondence may be addressed. Email: manthony5955@gmail.com or marcel.vanderheijden@agroscope.admin.ch.
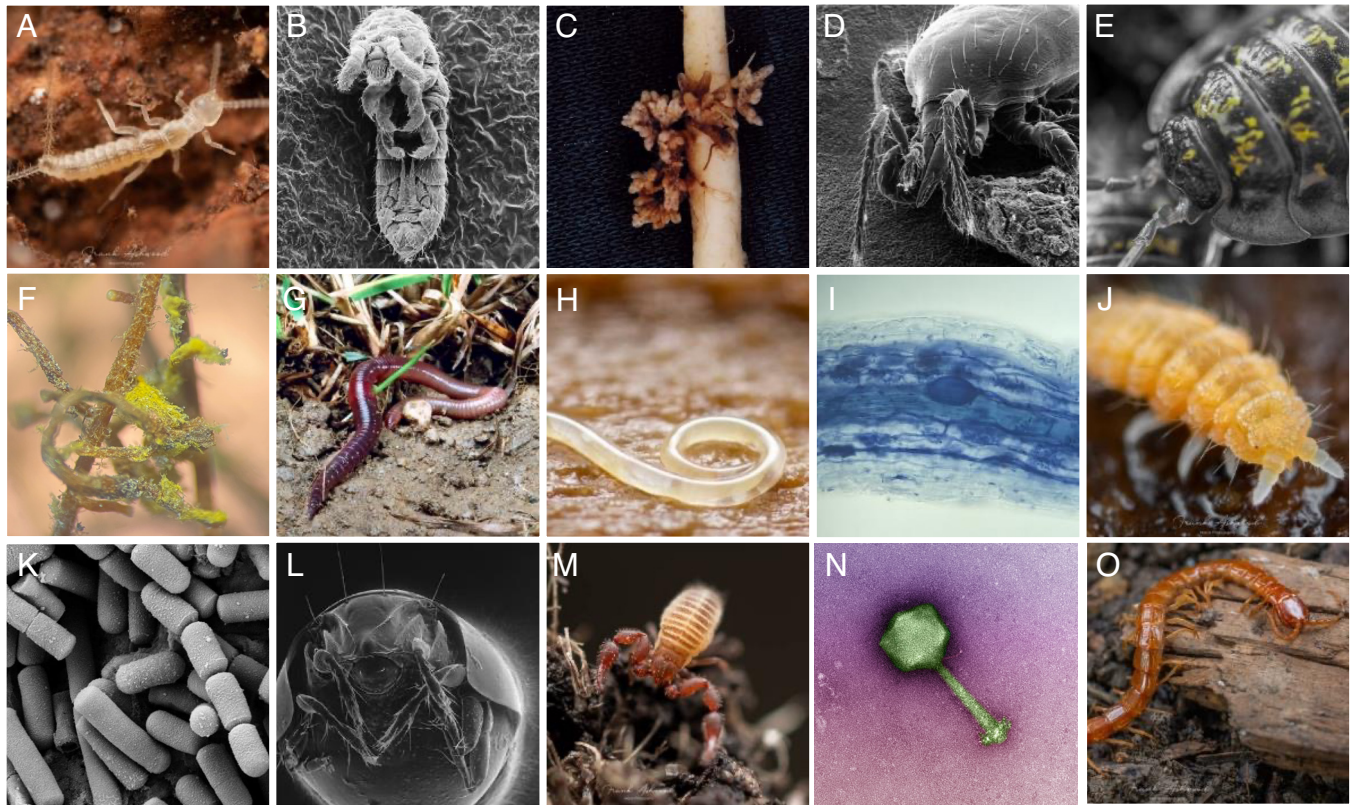
Fig. 1. Diversity of the major life forms found in soil. (*A*) bristletail (© F. Ashwood), (*B*) springtail (© H. Conrad), (*C*) nitrogen-fixing bacteria-containing nodules on clover root (© M. van der Heijden), (*D*) predatory mite (© H. Conrad), (*E*) isopod (© F. Ashwood), (*F*) scots pine root colonized by ectomycorrhizal fungi (yellow) (© M. Anthony), (*G*) earthworm (© G. Brändle), (*H*) nematode (© A. Murray), (*I*) corn root colonized by arbuscular mycorrhizal fungi (blue) (© F. Bender), (*J*) springtail (© F. Ashwood), (*K*) a common soil bacterium *Bacillus* (Creative Commons Attribution-Share license, photo by M. Das Murtey and P. Ramasamy), (*L*) horned mite (© H. Conrad), (*M*), pseudoscorpion (© F. Ashwood), (*N*) phage infecting a soil bacterium (© T. de Carvalho), (*O*) centipede (© F. Ashwood).

concept is even muddier for viruses due to immense ranges in mutation rates, lack of ribosomal and other replication-related protein-encoding genes, and the absence of reproduction (26). Yet, recent work demonstrates that viruses can be organized into isolated groups akin to the biological species concept (26), and overlapping viral RNA and DNA sequences can be used as indicators of taxa much like other microbial ribosomal RNA gene regions (27, 28). However, it is worth noting that the viral species concept is widely debated for myriad reasons (e.g., some viruses lack the ability to recombine genomes, viruses cannot replicate without hosts), and arguably, the viral species concept is most useful for taxonomic purposes (29, 30). Thus, while the species concept may vary considerably across groups, especially between macrobes and various microbes, there are now established methods and sufficient data to synthesize findings using best practices specific to each area of life.

However, there is one additional, large obstacle: We cannot rely on species counted by observation alone to accurately quantify total species diversity. If we relied on known species for all groups, like the previous estimate of soil animal diversity (12), the vast majority of small and cryptic lifeforms (e.g., Arthropoda, protists, Nematoda, Fungi, Bacteria, Archaea, Viruses) would not be counted (31–34). This is because for many groups, it is untenable to count all species, so we must use statistical approaches to more accurately approximate total species richness. Methods of prediction are diverse and include extrapolating proportions of known to unknown groups [e.g., plant:fungus ratios (35)], inference from rarefaction (36) or species-accumulation curves (37), estimators based on information of the rarest species [e.g., Chao and ACE

estimators (38)] or the least and most abundant species under a species abundance curve (39), extrapolations using species discoveries over time or estimates based on body size and year of description (40), or predictions based on known information from higher taxonomic levels (41). While each technique has its own set of assumptions and limitations, they have been employed at varying degrees across all major organismal groups ranging from microbes (34) to mammals (41) and provide invaluable and, oftentimes, more plausible enumeration of species numbers compared to direct counts alone.

Here, we review the literature of global species predictions to identify the lower, central, and upper estimations of global and soil-specific species richness on Earth. We include estimates using best practices respective to each organismal group, many based on statistical predictions of total species richness using theoretically grounded models. The central estimate is always in between the lower and upper and is generally the most rigorous richness estimate across multiple studies or based on the most comprehensive study to date. In addition to providing estimates of total numbers of species from the simplest (e.g., viruses) to most complex organisms (e.g., mammals), we also provide an estimate and error range of the proportion of life inhabiting soil (*Materials and Methods*). We emphasize that this effort is not encyclopedic, and it is not necessary to consider every group of organisms in order to estimate global biodiversity (*sensu* refs. 42 and 43). We included the most speciose groups and significant groups based on previous analyses and conservation efforts. The current state of soil biodiversity research is highly fragmented and speculative. Estimates provided here are therefore an initial attempt to unite disparate efforts and

highlight key knowledge gaps for the future. Previous reports of soil biodiversity have not been this expansive (44), nor have they differentiated between total versus soil-inhabiting species within groups (7, 45), including in the Global Soil Biodiversity Atlas (46). Our results therefore provide an estimate of biodiversity to enumerate the species richness of explicitly soil-inhabiting species. We define species that live in soil as those that live within, on (e.g., insects that feed on the surface of soil), or which complete any part of their life cycle in soil (e.g., organisms with an inactive pupal stage in soil or plant seeds that germinate in soil) or in the tissues of soil-dwelling symbionts (e.g., microbial parasites of soil animals). We keep a broad definition because these organisms rely on soil to grow, compete, and complete their life cycle. We define soil as the gases, solids, and water contained within the loose terrestrial surface covering Earth's upper crust. Whether viruses should be considered species has been questioned (see above), and because diversity estimates for viruses are also highly uncertain (see below), we provide estimates of soil biodiversity with and without viruses.

Results and Discussion

We first provide estimates of global macrobe species diversity with trackable estimates in soil for the most speciose groups based on literature review and data synthesis. Arthropoda are the most diverse group on Earth with a central estimate of 7×10^6 species (32) and a range of 1.14 to 30×10^6 (Table 1). In soil, we predict 2.1×10^6 species with a low of 3.49×10^4 and an upper prediction of 9.2×10^6 species (see estimate calculation and references in Table 1). We therefore estimate that $30.4 \pm 0.4\%$ of Arthropoda live in soil, with Isoptera ($84 \pm 15\%$ of species) and Diplopoda ($76.8 \pm 0.14\%$) being the most soil-specialized subgroups. On the other extreme are the Annelids, including the Enchytraeidae and Oligochaeta, with the lowest overall biodiversity but high specializations to soil. We estimate that there are 7.8×10^2 and 1×10^3 Enchytraeidae and Oligochaeta species and that $98.6 \pm 0.06\%$ and $63 \pm 4.2\%$ of species live in soil, respectively. The least speciose group in soil are the Mammalia (3.8% of species, no error range estimated) and Mollusca ($20.4 \pm 17.6\%$). Neither of these results is surprising since Mammalia and Mollusca lifeforms have primarily evolved to live aboveground (47) and in the oceans (48), respectively.

Most Plantae can be considered to live in soil because of their belowground root systems and life cycles in soil. Between 22 and 67% of terrestrial plant biomass is in soil (i.e., roots) (79), and many terrestrial plant seeds and spores germinate in soil. Current best estimates suggest that there are 5.37×10^5 plant species (Table 1, central estimate). Approximately 7.2 to 11.4% of angiosperm families are entirely or partially aquatic (80), equivalent to about ca. 39,453 species. There are also ca. 31,311 vascular epiphytes (81). After accounting for epiphytes and aquatic angiosperms, there are ca. 4.7×10^5 plant species that live in soil. Incorporating all estimates of plant diversity together, we estimate that $85 \pm 1.3\%$ of plant species live in soil.

Estimates of biodiversity become more complicated for smaller organisms like Nematoda. The current best estimate suggests that there are 1×10^6 Nematoda species, defined based on rRNA gene dissimilarity, morphological-based surveys, and statistical inference (Table 1, central estimate), but estimates range from an unrealistic low of 2×10^4 species (7) to as high as 1×10^8 species (52). We can reject this low estimate since there are already 2.7×10^4 known Nematoda species (82) and embrace a more realistic low estimate of 5×10^5 species (51). Bagyaraj et al. (50) argue that half of all Nematoda species live in soil as plant and animal parasites or as free-living species, based on expert opinion. The share of Nematoda living as helminths inside soil animals represents a large fraction of this biodiversity (42, 82), and we include them here as soil-dwelling to be consistent with earlier definitions (50, 83). A synthesis of marine biodiversity shows 1.1×10^4 described Nematoda species and an estimated 5×10^4 or more species yet to be discovered (84). There is no global estimate of fresh-water Nematoda biodiversity, but 7% of the currently described 2.7×10^4 species live in fresh-water (85), and 50% live in the marine environment (50). If we assume that this 57% reflects nonsoil inhabiting species, then we can estimate that 43% of Nematoda biodiversity lives in soil (or in animals and plants that live in soil), a value not far from the expert opinion of Bagyaraj et al. (50). If there are 1×10^6 Nematoda species, we estimate that $4.3\text{-}5 \times 10^5$ species live in soil.

Protists are polyphyletic, microscopic eukaryotes that are not Animalia, Plantae, or Fungi. There is a massive range in global protist diversity from 6×10^4 up to 16.2×10^7 with a central estimate of 2×10^5 (Table 1). The upper prediction is orders of magnitude higher than the central prediction, and it is based on ratios of protists to Arthropoda (42). While possible, predictions of biodiversity using the protist to Arthropoda ratio have been highly criticized in the literature (19), so we conservatively do not consider them when calculating soil protist diversity. Despite these concerns, it is noteworthy that tropical rainforest soil protist communities are mainly composed of host-specific parasitic species (83), of which many are linked to Arthropoda. Further studies are necessary to assess whether the protist to Arthropoda ratio is a useful proxy to estimate protist diversity. To estimate the proportion of protists in soil, we calculated the percent of species observed in soil relative to other ecosystem types using two independent, global datasets. We found that 40.4% of amplicon sequence variants (ASVs), a molecular metric for an individual taxon, were detected in soil compared to city water, lakes, and marine ecosystems (86) (*SI Appendix*, Table S1). In the second dataset, we found that 42.5% of operational taxonomic units (OTUs), another molecular species metric, were detected in soil compared to freshwater and marine ecosystems (87) (*SI Appendix*, Table S2). Estimates for the fraction of taxa living in soil are therefore very similar when using ASVs or OTUs as species metrics, though it is important to acknowledge that these are likely underestimates since most widely used DNA metabarcoding and bioinformatic techniques omit many soil protists known to be abundant based on morphological observation (88). We therefore cautiously use the lower estimate of the percent of species found in soil (40.4%) and the average of lower and higher estimates (41.45%) to compute the fraction of protist species found in soil relative to global predictions, accepting that this is likely an underestimation. We estimate that $2.4\text{-}8.3 \times 10^4$ protist species live in soil, and we use this number because it reflects the range in estimates for the fraction of taxa found in soil relative to global predictions.

The other large group of microscopic eukaryotes is Fungi, whose soil biodiversity has received considerable attention with numerous global biodiversity surveys (69, 89–91). Fungi are one of the most biodiverse groups of organisms on Earth with global estimates as low as 2.2×10^6 and as high as 1.65×10^8 (Table 1). The current best estimate, from a recent compilation of ITS2 ribosomal RNA gene sequences in the GlobalFungi database, predicts 6.2×10^6 species computed as 97% sequence similarity OTUs (69). This number is likely even higher because many species within large fungal genera are combined into one taxon at 97% sequence similarity (24), and sampling to date is highly skewed toward northern latitudes (92). The upper estimate of fungal diversity (1.65×10^8) has the same limitations as the upper protists estimate—it is based

Table 1. Enumeration of biodiversity on Earth and in soil

Macrobes	Lower	Central	Upper	Conf.
		Total species		
Mammalia	n/a	6.5×10^3 (16, 49)	4.5×10^4 (41)	H
Nematoda	5×10^5 (50, 51)	1×10^6 (33, 52)	1×10^8 (52)	L
Arthropoda	1.14×10^6 (53)	7×10^6 (32)	3×10^7 (54)[b,?]	M
Insecta	9.5×10^5 (53)	$5\text{-}5.5 \times 10^6$ (32, 41)	1×10^8 (55)[b,?]	M
Arachnida	9×10^4 (53)	1.1×10^5 (41)	7.5×10^5 (50)	M
Collembola	6.5×10^3 (56)	8.7×10^3 (53)	5×10^4 (57)	M
Diplopoda	1.3×10^4 (53, 58)	1.5×10^4 (58)	$7\text{-}8 \times 10^4$ (41, 59)	M
Isoptera	3×10^3 (53)	5.4×10^3 (60)	1×10^4 (7)	M
Formicidae	1×10^4 (53)	1.5×10^4 (61)	2.1×10^4 (62)	M
Oligochaeta	5×10^3 (63)	1×10^4 (64)	?	H
Enchytraeidae	7.1×10^2 (65)	7.8×10^2 (53)	1.2×10^3 (7)	H
Mollusca	8×10^4 (41)	1.2×10^5 (53)	2×10^5 (50)	H
Plantae	3.8×10^5 (53)	5.37×10^5 (64)	n/a	H
Microbes[k]	$6.7 \times 10^{6,k}$	$1.01 \times 10^{9,k}$	$1 \times 10^{12\text{-}14}$ (34, 66)[?]	L
Phage	1×10^8 (67)[?]	1×10^{11} (68)[l,?]	3.7×10^{11} (68)[l,?]	VL
Bacteria	$4.3\text{-}4.5 \times 10^6$ (19, 39)	1×10^9 (17)	3.7×10^9 (43)[n]	L
Fungi	2.2×10^6 (35)	6.2×10^6 (69)	1.65×10^8 (42)[p,?]	M
Archaea	1.1×10^4 (70)	1.9×10^5 (19)	3.8×10^5 (19)[r]	L
Protists	6×10^4 (55)	$1\text{-}3 \times 10^5$ (12, 55, 71)	$8.4\text{-}16.2 \times 10^7$ (42, 72)[?]	L
Total	1.1×10^8	1.01×10^{11}	3.74×10^{11}	L
Macrobes		Species in soil		
Mammalia	75 (12)	250 (73)	?	H
Nematoda	$2.15 \times 10^{5,a}$	$4.65 \times 10^{5,a}$	$5 \times 10^{7,a}$	L
Arthropoda	3.49×10^5 (12)	$2.1 \times 10^{6,c}$	$9.2 \times 10^{6,c}$	M
Insecta	2.88×10^5 (12)	$1.67 \times 10^{6,c}$	$3 \times 10^{6,c}$	M
Arachnida	4.32×10^4 (12)	$5 \times 10^{4,d}$	$3.5 \times 10^{5,c}$	M
Collembola	5.98×10^3 (74)[e]	8×10^3 (12)	8.18×10^3 (74)[e]	M
Diplopoda	1×10^4 (12)	$1.15 \times 10^{4,c}$	$6.15 \times 10^{4,c}$	M
Isoptera	2×10^3 (12)	$4.9 \times 10^{3,f}$	$9.5 \times 10^{3,f}$	M
Formicidae	5×10^3 (75)[g]	9×10^3 (12)	1.1×10^4 (75)[g]	M
Oligochaeta	3.3×10^3 (63)	6×10^3 (12)	?	H
Enchytraeidae	$7 \times 10^{2,h}$	7.7×10^2 (12)	$1.18 \times 10^{3,c}$	H
Mollusca	80 (12)	3.5×10^4 (76)[i]	6.4×10^4 (77)[i]	H
Plantae	$3.2 \times 10^{5,j}$	$4.66 \times 10^{5,j}$	n/a	H
Microbes[k]	$6 \times 10^{6,k}$	$4.4 \times 10^{8,k}$	?	L
Phage	5.6×10^6 (68)[m,?]	9.9×10^9 (78)[m,?]	$1.59 \times 10^{11,m,?}$	VL
Bacteria	9.5×10^5 (39)[o]	$4.3 \times 10^{8,o}$	$3.3 \times 10^{9,o}$	L
Fungi	$2 \times 10^{6,q}$	$5.6 \times 10^{6,q}$	n/a[q]	M
Archaea	$1.1 \times 10^{3,s}$	$3.6 \times 10^{4,s}$	$1.9 \times 10^{5,s}$	L
Protists	$2.4 \times 10^{4,t}$	$8.3 \times 10^{4,t}$	n/a[t]	L
Total	9.5×10^6	1.04×10^{10}	1.62×10^{11}	L

Species in soil include taxa that live in, on, or which complete part of their life cycle in soil. Values representing the most realistic predictions at the lower, central, and upper ranges are included based on literature review. For microbes, species concepts are based on direct counts or predictions of the most accepted molecular definitions of "species" (unless explicitly stated elsewise) represented as contigs, protein clusters, or viral OTUs for viruses, and 97% sequence similarity OTUs for protists, bacteria, archaea, and fungi. Note that each reference is shown in parentheses as a numeric value whereas methodological notes describing the estimations are shown outside of the parentheses using letters (*SI Appendix*, Table S10), where needed. Groups within a larger category are shown in italicized font. Question marks denote areas where we currently lack enough information to estimate. Not applicable (n/a) refers to lower and/or upper estimations that are obsolete or which do not make sense to estimate because there is no realistic lower or upper prediction due to a high confidence central prediction. In the absence of a value or where a value is highly speculative (e.g., upper fungal diversity), we used the central estimate to predict total species richness in soil. The confidence column (Conf.) reflects how well current data support predictions of species in soil within each group as very low (VL), low (L), mid (M), or high (H).

on ratios of Fungi to Arthropoda (42) and was therefore not included when predicting soil fungal biodiversity. Global surveys suggest that there are 3×10^3 lichenicolous fungi, 8.4×10^3 aquatic fungi, 2×10^4 Arthropoda-associated fungi, and 1.2×10^3 microsporidians (mostly animal parasites)[4], equivalent to ca. 3.3×10^4 potentially non-soil-dwelling fungal species. We also estimate fungal diversity in animal hosts to be at least 10× lower than soil fungal diversity (93). Thus, if we account for the 3.3×10^4 non-soil-dwelling fungi, we estimate that 10% of the remaining fungal diversity is animal associated. Assuming that these fungi

generally lack dual capacities to live biotrophically in animal hosts and as soil saprotrophs, an assumption that we know is not correct for all species (94), we estimate between 1.9 and 5.99×10^5 nonsoil inhabiting fungi and 2 and 5.6×10^6 species of soil-dwelling fungi. This estimate would be about 25% higher if ASVs were used as species proxies versus the 97% OTU threshold used here (95), but this approach may not be advisable for fungal ITS-based surveys where we observe sequence variability within species for some fungal clades (24). Soil fungal diversity is therefore one of the highest compared to global diversity among all groups of life on Earth.

Bacteria are an order of magnitude more diverse than Fungi and Nematoda combined, but their soil biodiversity is harder to estimate. Global estimates of bacterial richness range from 4.3×10^6 to 3.7×10^9 with a central estimate of 1×10^9 (Table 1). To derive estimates of soil bacterial richness, we reanalyzed three global prokaryotic datasets to estimate the proportion of observed bacterial species in soil versus other ecosystem types. We first reanalyzed data from the SILVA database —a quality-checked resource of ribosomal RNA gene sequences (96) with metadata organized by others (70)—to estimate that 27.9% of all 97% OTUs have been observed in soil, after correcting for sequencing effort (*SI Appendix,* Table S3). We then repeated this process using data from the Earth Microbiome Project (97) and found that 32.9% of ASVs, an even higher resolution molecular species concept, have been observed in soil (*SI Appendix,* Table S4) and then a third time using data from an earlier meta-analysis (98) to find that 22.2% of OTUs have been observed in soil (*SI Appendix,* Table S5). All three estimates are substantially lower than an initial estimate of 88.8% calculated using theoretical statistical approaches (39). Estimates of the proportion of bacterial life in soil therefore range between 22 and 89% with an average (hereafter: central estimate) of 43%. Using 22.2, 43, and 88.8% as our low, central, and upper predictions of the proportion of bacterial species in soil, we estimate that there are 9.5×10^5, 4.3×10^8, and 3.3×10^9 bacterial species in soil, respectively. We used the same methodology as described above to predict archaeal soil richness. Global estimates of archaeal richness are 1.1×10^4, 1.9×10^5, and 3.8×10^5 at the lower, central, and upper levels of prediction. Using two independent datasets, we calculate the proportion of Archaea OTUs in soil to range from 10.3 to 27.8% with an average of 19.1% (*SI Appendix,* Tables S6 and S7). Using 10.3, 19.1, and 27.8% as our lower, central, and upper soil proportion estimates, we predict that there are 1.1×10^3, 3.6×10^4, and 1.1×10^5 Archaeal OTUs in soil, respectively.

It is worth noting that the values for prokaryotes outlined above are potentially considerable underestimations. There are two estimates of global microbial richness that suggest orders of magnitude higher diversity ($1 \times 10^{12-14}$) (34, 66). These staggering estimates include all microbial groups (except protists) but primarily represent bacteria and are computed using power-scaling laws and global compilations of 16S ribosomal RNA sequences clustered at a 97% sequence similarity. While we did not include these values in our ultimate predictions of soil biodiversity, it is possible that bacterial richness is orders of magnitude higher than reported here. And, if ASVs were used, we would expect these values to be twice as high (95) and 2.7 times higher if full-length 16S versus shorter hypervariable regions were analyzed (19). Even if the exact species richness estimates presented here are underestimates, the proportions of species living in soil would remain unaffected. Soil is therefore a massive reservoir of bacterial species richness consistent with general soil biodiversity theory. Our estimate of the proportion of life in soil is also higher than any other singular ecosystem type where we also observe high bacterial species richness, including in sediments, air, water, and host organisms such as arthropods and humans (*SI Appendix,* Table S3–S5).

The most challenging groups for which we estimated biodiversity are viruses. A major reason is that viral and bacterial diversity are not independent. Estimates of viral biodiversity focus on phages (viruses that infect bacteria) because bacteria are orders of magnitude more diverse than any other group (Table 1). The first step to quantifying viral biodiversity is therefore accurately identifying how many unique phages are typically found per bacterial species. Traditional models suggest a virus-to-bacterium ratio of 10:1 (99), but this is likely a gross underestimation. Thirty distinct viral sequence clusters, a molecular indicator of an individual viral taxon, have been isolated from hundreds of *Mycobacterium smegmatis* cultures (100), and many of the clusters are so diverse that they likely reflect more than one taxon. Hundreds to thousands of viral protein clusters, another molecular metric for a viral taxon, have been discovered in the pangenomes of individual bacterial species of *Pseudomonas*, *Straphylococcus*, *Mycobacterium*, and *Lactococcus* (68). There are more likely hundreds to thousands of unique phage taxa infecting any individual bacterial species. Phage also typically infect only a single host species or strain (101) because each host harbors a unique repertoire of cell-binding surface receptors that the phage must recognize (102). By analyzing co-occurrence patterns of viral OTUs from bacterial hosts with at least 1,000 viral metagenomes in the IMG/VR database (78), we show that phages are rather host specific and that bacterial hosts generally share 0 to 1 viral taxa (*SI Appendix,* Fig. S1A), though as many as 53 vOTUs were shared between *Escherichia coli* and *Salmonella enterica*. Each of these bacterial species also harbored hundreds to thousands of vOTUs. We can therefore use the virus-to-bacterium ratio as a method to predict phage species richness assuming minimal shared viral species among distinct hosts (68). If we assume an intermediate virus-to-bacterium ratio of 1,000:1 to estimate phage diversity, we predict anywhere from 1×10^8 to 3.7×10^{11} viral "species". This upper range is much higher than previous global phage diversity estimates of 1 to 6 billion (67, 68) because we used more accurate, updated predictions of global bacterial species richness. Even if we used a much more conservative estimate of the virus-to-bacterium ratio of 100:1, we would still predict upward of 1×10^{10} phage species.

Next, we incorporated our predictions of global phage biodiversity with data from global surveys of viral diversity across ecosystem types to provide an initial estimate of soil viral biodiversity. To do this, we downloaded the most recent and largest dataset of viral metagenomes from IMG/VR (78), and we compared the distribution of vOTU richness across ecosystem types to date (*SI Appendix,* Fig. S1B). We found that 9.9% of all vOTUs were from soil (*SI Appendix,* Fig. S1C). We arrived at a slightly smaller estimate of 5.6% using an independent dataset of 1,882 metagenomic samples (*SI Appendix,* Table S8). We therefore estimate that a minimum of 6 to 10% of viral diversity is found in soil and acknowledge that this value is much likely higher since the databases we used, albeit the best ones currently publicly available, are highly biased toward oceans and host associates (humans) (*SI Appendix,* Table S8). For example, a comparison of two fresh-water lakes to agricultural soils collected around the catchment zone found 50 to 1,000% higher viral richness in soil compared to lakes (103). Soil viral communities also exhibit remarkable distance decay, with ~75% of vOTUs different across a 10-m distance (104), consistent with other studies showing that >90% of vOTUs at a given site are unique (105, 106). Viral species composition is also tightly linked to bacterial species composition (103, 104), though both spatial (104) and temporal (107) turnover is greater for soil viruses

compared to bacteria. Because most viruses are phages, we expect the proportion of phages in soil to track soil bacterial diversity, with the possibility for soil phage biodiversity to be higher due to the aforementioned higher turnover rates. Our central estimate for the proportion of bacterial life in soil is 43%, and if we substitute this value as a theoretical estimate of phage soil biodiversity, we estimate that a range of 5.6×10^6, 9.9×10^9, and 1.59×10^{11} viral taxa live in soil, assuming a 6% lower, 9.9% middle, and 43% upper estimates (average of $19.5 \pm 12\%$) of the proportion of viral life in soil. When we consider our estimates in the context of the current state of soil virology research, soil is a major reservoir of viral biodiversity at the global scale. Moving forward, it will be important for more concerted efforts to sample soil viral metagenomes across the globe and to apply complementary tools like electron microscopy and fluorescent staining to attempt to differentiate between DNA and RNA viruses (108).

When considering most life on Earth together, the average proportion of species in soil across all three estimates (lower, central, and upper) is $58.5 \pm 14.7\%$, excluding phage (Figs. 2 and 3). Because phage biodiversity is orders of magnitude higher than for all other groups, a soil estimate including phages (20.8 ± 11) is essentially the same as the average proportion of phages in soil (19.5 ± 12) and is therefore not representative of most groups of life. Our prediction of the proportion of species in soil is not strongly affected if we also remove bacteria alongside phages, the second most speciose group (63 ± 5 versus $58.5 \pm 14.7\%$). However, the total number of species in soil is estimated to be between 9.5×10^6 and 1.62×10^{11}, and this number is strongly affected the most species-rich groups like bacteria but especially phages. Since we have the lowest confidence in our phage estimates (Table 1), and the viral species concept is arguably a more utilitarian than genetic concept, there is greater value to our estimate reflecting life beyond phages. This value better represents the proportion of life in soil across the tree of life. Even if we only considered the proportion of life based on the central prediction of species richness (43.2%), this is still a massive fraction of the total proportion of life on Earth.

We therefore estimate more than two times greater life in soil than former predictions by Decaëns et al. (12). Our estimate that an average of 58.5% of life inhabits soil provides a staggering enumeration to the value of conserving soil biodiversity. While some predictions range across orders of magnitude (e.g., Bacteria, phage) and are currently limited by data availability and potential shortcomings of the models used to predict species richness (e.g., protists, phage), we at least suggest that the current estimate of 25% strongly underestimates soil biodiversity. Moving forward, we encourage particularly greater attention to biodiversity exploration for viruses, bacteria, and protists. Most large-scale efforts to sample viruses have been in the global oceans (109), and there is a massive need for viral discovery in soils (110, 111). There is also a sizable range in the bacterial predictions presented here that may be biased by massive efforts to sample the human microbiome relative to soils (70) (*SI Appendix*, Table S1). This same work shows that the number of newly detected bacterial taxa is decreasing over time (70). This is a result of geographic undersampling of many areas of the world (112), including in often-neglected deep soils which harbor many unique lineages compared to surface soil (113). This emphasizes the need to orchestrate deliberate soil sampling efforts toward areas not yet captured by existing observations or environmental conditions (92), but this effort should not stop with microorganisms. In fact, our predictions of the proportion of species in soil is likely underestimated for many groups because Earth's life has been so much more extensively characterized in above-ground, nonsoil systems across the tree of life.

Our goal with this effort was to provide an estimate of the proportion of life inhabiting soil. Because soil organisms drive global biogeochemical cycles and human health and form the base of agricultural food and fiber production, they are indispensable components of the biosphere. Many ecosystems across the globe are not considered in conservation efforts owing to a lack of visible organisms of interest (114). This inherently overlooks many soil organisms whose cryptic lifestyles are mostly incompatible with existing structures for the conservation of species. Changing this requires greater attention and advocation by the public, scientists, and policy makers for soil biodiversity. We suggest that a value of
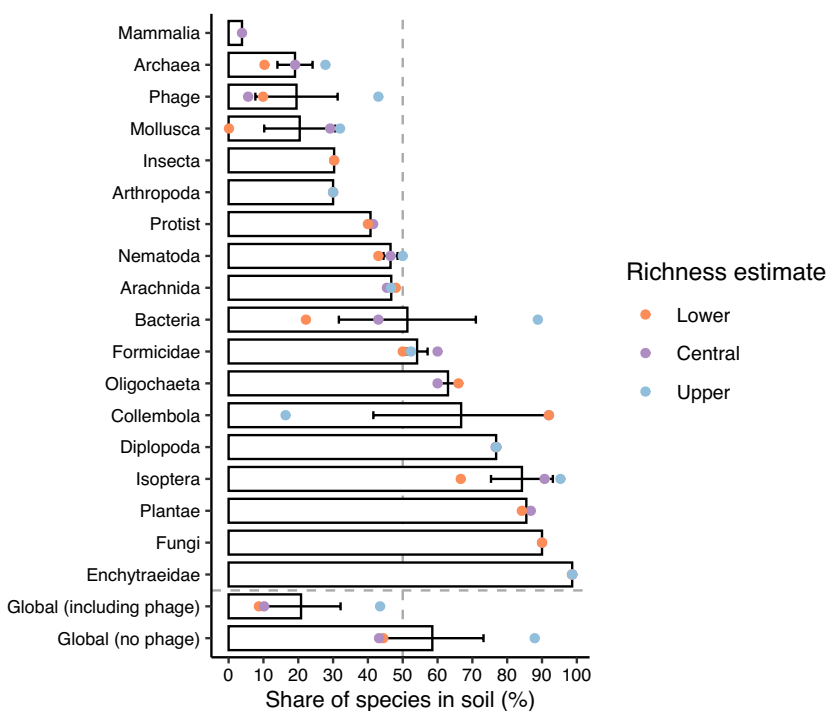
Fig. 2. Percentages of different groups of life on Earth that live in soil. Species that live in, on, or complete part of their life cycle in soil were classified as "species in soil". The global percent of species in soil, including or excluding phage, summarized across all groups is also shown. Bars represent the mean and error bars are plus or minus one SE (where >1 prediction was available) of the lower, central, and upper estimates (e.g., for bacteria, we estimate 22.2, 43, and 88.8% for the lower, central, and upper estimates, respectively, for an average of 51 ± 20%). Note that some groups do not have a lower, central, and upper species richness estimate (e.g., Mammalia; see Table 1), and there are no error bars for these groups. See Table 1 for an estimate of the total number of taxa on Earth and those living in soil for each taxonomic group as well as our confidence in each prediction.
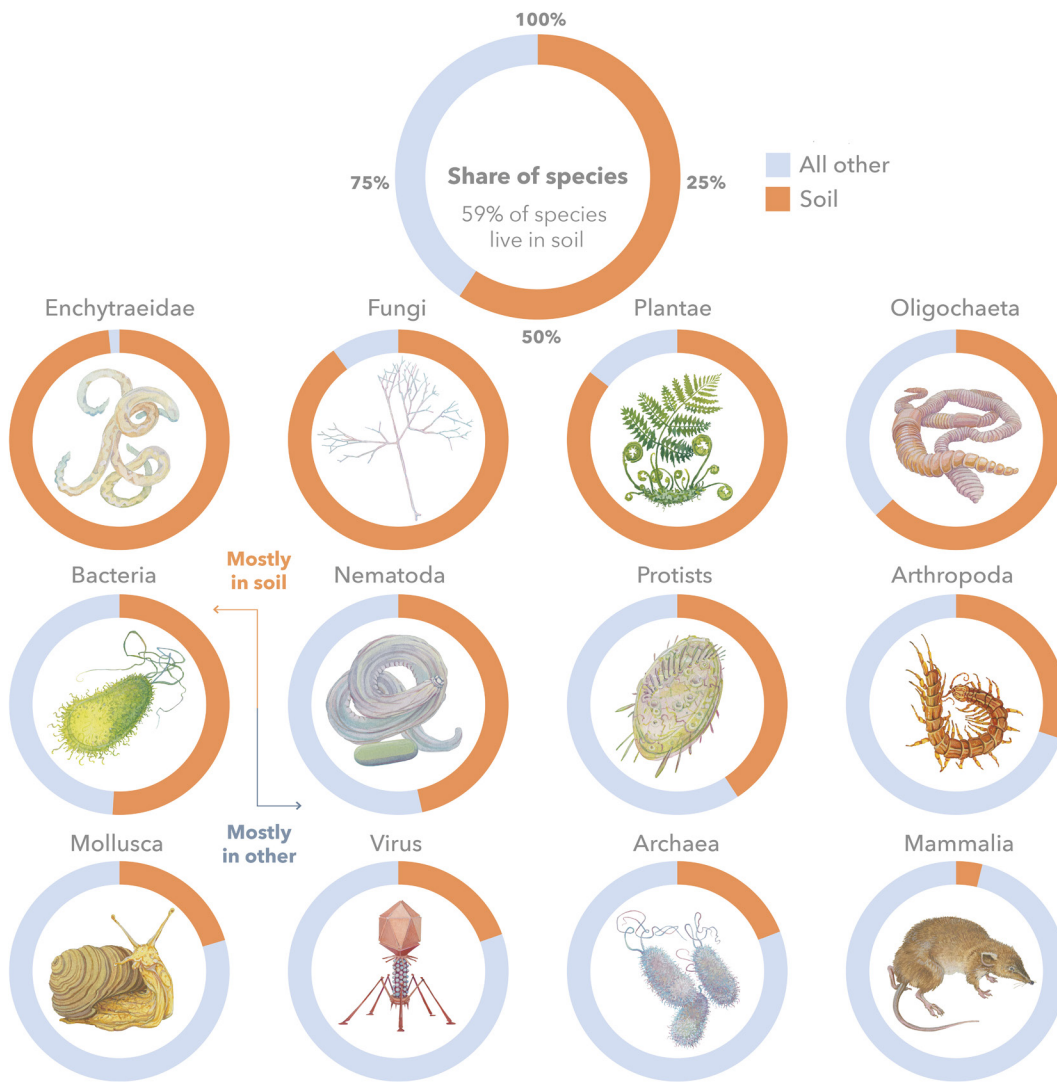
Fig. 3. Graphical overview of the share of species living in soil. Doughnuts reflect the percentage of species in soil versus all other ecosystems combined (e.g., marine, freshwater, built environment, host organisms such as humans). The larger doughnut on top shows the total share of species, and smaller doughnuts show individual shares for the most speciose and well-known groups ordered from greatest to least specialized in soil. Illustrations by Michael Dandley ©.

59% be used to advocate for soil biodiversity conservation and restoration as a central goal of the Anthropocene.

Materials and Methods

We searched the literature for estimates of total species richness for most groups of life (see a list of groups in Table 1) using Google Scholar. We searched for *biodiversity, *diversity, *species richness, *global, and *total species (where * represents search terms with and without the word "soil") and browsed article titles, abstracts, and figures to deduce whether there was quantitative information that could be included as an estimate of species richness. We conducted the same searches for each organismal group listed in Table 1.

Methods of quantifying species biodiversity and the number of publications differed widely across groups considered in this study. We therefore employed a variety of complimentary techniques specific to certain groups to estimate soil biodiversity (see Methodological notes for in *SI Appendix*, Table S10 for a comprehensive overview). When a direct soil estimation was unavailable or unable to be estimated using the current literature, we downloaded global datasets of species distributions for individual groups (e.g., viruses, bacteria, archaea, protists) and calculated the share of observed species in soil relative to other ecosystem types. We then used this proportion to estimate the share of species in soil based on the total species estimate.

When possible, we provide three feasible estimates of species biodiversity: a lower, central, and upper prediction. We then used these estimates to calculate the proportion of species in soil using the following equation:

$$\text{Proportion of species in soil} = \bar{x}\left(\left(\sum \text{soil species (upper)} \div \sum \text{total species (upper)}\right)\right.$$
$$+ \left(\sum \text{soil species (central)} \div \sum \text{total species (central)}\right)$$
$$\left.+ \left(\sum \text{soil species (lower)} \div \sum \text{total species (lower)}\right)\right).$$

1. A. M. Elewa, "Mass extinction-a general view" in *Mass Extinction* (Springer, 2008), pp. 1–4.
2. D. B. Wake, V. T. Vredenburg, Are we in the midst of the sixth mass extinction? A view from the world of amphibians *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11466–11473 (2008).
3. D. S. Wilcove, D. Rothstein, J. Dubow, A. Phillips, E. Losos, Quantifying threats to imperiled species in the United States: Assessing the relative importance of habitat destruction, alien species, pollution, overexploitation, and disease. *BioScience* **48**, 607–615 (1998).
4. J. Gurevitch, D. K. Padilla, Are invasive species a major cause of extinctions? *Trends Ecol. Evol.* **19**, 470–474 (2004).
5. R. D. Bardgett, W. H. van der Putten, Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
6. M. G. A. Van Der Heijden, R. D. Bardgett, N. M. Van Straalen, The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **11**, 296–310 (2008).
7. U. N. Nielsen, D. H. Wall, J. Six, Soil biodiversity and the environment. *Annu. Rev. Environ. Resour.* **40**, 63–90 (2015).
8. S. Banerjee, M. G. A. van der Heijden, Soil microbiomes and one health. *Nat. Rev. Microbiol.* **21**, 6–20 (2023).
9. M. A. Anthony et al., Forest tree growth is linked to mycorrhizal fungal composition and function across Europe. *ISME J.* **16**, 1327–1336 (2022).
10. T. W. Crowther et al., The global soil community and its influence on biogeochemistry. *Science* **365**, eaav0550 (2019).
11. E. Ladouceur et al., Knowledge sharing for shared success in the decade on ecosystem restoration. *Ecol. Solutions Evidence* **3**, e12117 (2022).
12. T. Decaëns, J. J. Jiménez, C. Gioia, G. Measey, P. Lavelle, The values of soil animals for conservation biology. *Eur. J. Soil Biol.* **42**, S23–S38 (2006).
13. E. Mayr, *Systematics and the Origin of Species* (Columbia University Press, New York, 1942), p. 334.
14. F. E. Zachos, L. Christidis, S. T. Garnett, Mammalian species and the twofold nature of taxonomy: A comment on Taylor et al. 2019. *Mammalia* **84**, 1–5 (2020).
15. J. W. Sites, J. C. Marshall, Operational criteria for delimiting species. *Annu. Rev. Ecol. Evol. Syst.* **35**, 199–227 (2004).
16. C. J. Burgin, J. P. Colella, P. L. Kahn, N. S. Upham, How many species of mammals are there? *J. Mammal.* **99**, 1–14 (2018).
17. D. Dykhuizen, Species numbers in bacteria. *Proc. Calif. Acad. Sci.* **56**, 62 (2005).
18. P. D. Schloss, Amplicon sequence variants artificially split bacterial genomes into separate clusters. *Msphere* **6**, e00191-21 (2021).
19. S. Louca, F. Mazel, M. Doebeli, L. W. Parfrey, A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS Biol.* **17**, e3000106 (2019).
20. C. M. Thomas, K. M. Nielsen, Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
21. N. Corradi, A. Brachmann, Fungal mating in the most widespread plant symbionts? *Trends Plant Sci.* **22**, 175–183 (2017).
22. J. M. Janda, S. L. Abbott, 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J. Clin. Microbiol.* **45**, 2761–2764 (2007).
23. E. Stackebrandt, B. M. Goebel, Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Intern. J. Syst. Evol. Microbiol.* **44**, 846–849 (1994).
24. L. Tedersoo et al., Best practices in metabarcoding of fungi: From experimental design to results. *Mol. Ecol.* **31**, 2769–2795 (2022).
25. K. S. Ramirez et al., Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat. Microbiol.* **3**, 189–196 (2018).
26. L.-M. Bobay, H. Ochman, Biological species in the viral world. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6040–6045 (2018).
27. S. Roux et al., Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).
28. S. Roux et al., Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37 (2019).
29. M. H. Van Regenmortel, Viruses are real, virus species are man-made, taxonomic constructions. *Arch. Virol.* **148**, 2481 (2003).
30. F. M. Zerbini et al., Differentiating between viruses and virus species by writing their names correctly. *Arch. Virol.* **167**, 1231–1234 (2022).
31. M. Achtman, M. Wagner, Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* **6**, 431–440 (2008).
32. N. E. Stork, How many species of insects and other terrestrial arthropods are there on Earth? *Annu. Rev. Entomol.* **63**, 31–45 (2018).
33. K. Kiontke, D. H. Fitch, Nematodes. *Curr. Biol.* **23**, R862–R864 (2013).
34. K. J. Locey, J. T. Lennon, Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5970–5975 (2016).
35. D. L. Hawksworth, R. Lücking, Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol. Spectr.* **5**, 5–4 (2017).
36. N. J. Gotelli, R. K. Colwell, Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**, 379–391 (2001).
37. K. I. Ugland, J. S. Gray, K. E. Ellingsen, The species–accumulation curve and estimation of species richness. *J. Animal Ecol.* **72**, 888–897 (2003).
38. B.-R. Kim et al., Deciphering diversity indices for a better understanding of microbial communities. *J. Microbiol. Biotechnol.* **27**, 2089–2093 (2017).
39. T. P. Curtis, W. T. Sloan, J. W. Scannell, Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10494–10499 (2002).
40. N. E. Stork, J. McBroom, C. Gely, A. J. Hamilton, New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7519–7523 (2015).
41. C. Mora, D. P. Tittensor, S. Adl, A. G. Simpson, B. Worm, How many species are there on Earth and in the ocean? *PLoS Biol.* **9**, e1001127 (2011).
42. B. B. Larsen, E. C. Miller, M. K. Rhodes, J. J. Wiens, Inordinate fondness multiplied and redistributed: The number of species on Earth and the new pie of life. *Q. Rev. Biol.* **92**, 229–265 (2017).
43. J. J. Wiens, Vast (but avoidable) underestimation of global biodiversity. *PLoS Biol.* **19**, e3001192 (2021).
44. S. Geisen et al., A methodological framework to embrace soil biodiversity. *Soil Biol. Biochem.* **136**, 107536 (2019).
45. M. Swift, D. E. Bignell, F. M. S. Moreira, E. Huising, "The inventory of soil biological diversity: Concepts and general guidelines" in *A Handbook of Tropical Soil Biology*, F. M. S. Moreira, E. J. Huising, D. E. Bignell, Eds. (Earthscan, London, 2008).
46. A. Orgiazzi, R. D. Bardgett, E. Barrios, *Global Soil Biodiversity Atlas* (European Commission, 2016).
47. P. Hunter, The rise of the mammals: Fossil discoveries combined with dating advances give insight into the great mammal expansion. *EMBO Rep.* **21**, e51617 (2020).
48. G. Vermeij, R. Dudley, Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? *Biol. J. Linnean Soc.* **70**, 541–554 (2000).
49. Mammal Diversity Database, Mammal diversity database (1.9) [Data set]. *Zenodo* (2022). https://doi.org/10.5281/zenodo.6407053 (21 October 2022).
50. D. Bagyaraj, C. Nethravathi, K. Nitin, "Soil biodiversity and arthropods: Role in soil fertility" in *Economic and Ecological Significance of Arthropods in Diversified Ecosystems* (Springer, 2016), pp. 17–51.
51. United Nations Environment Programme World Conservation Monitoring Centre (United Nations Environment Programme-WCMC), *Global biodiversity: Status of Earth's living resources* (Chapman and Hall, London, UK, 1992).
52. P. J. D. Lambshead, "Marine nematode biodiversity" in *Nematology: Advances and Perspectives. Volume 1: Nematode Morphology, Physiology, and Ecology* (CABI Books, CABI International, Wallingsform, UK, 2004), pp. 438–468.
53. O. Bánki et al., Catalogue of life checklist (2022), https://doi.org/10.48580/dfqc.
54. T. L. Erwin, Tropical forests: Their richness in Coleoptera and other arthropod species. *Coleopterists Bull.* **36**, 74–75 (1982).
55. K. J. Gaston, "Global species richness" in *Encyclopedia of Biodiversity* (Academic Press, San Diego, CA, 2008).
56. J. Rusek, Biodiversity of Collembola and their functional role in the ecosystem. *Biodiversity Conserv.* **7**, 1207–1219 (1998).
57. S. P. Hopkin, *Biology of the Springtails:(Insecta: Collembola)* (OUP Oxford, 1997).
58. M. S. Brewer, P. Sierwald, J. E. Bond, Millipede taxonomy after 250 years: Classification and taxonomic practices in a mega-diverse yet understudied arthropod group. *PLoS One* **7**, e37240 (2012).
59. R. Hoffman, *Classification of the Diplopoda* (Muséum d'Histoire Naturelle, Genève, 1980), pp. 1–237.
60. R. Constantino, Estimating global termite species richness using extrapolation. *Sociobiology* **65**, 10–14 (2018).
61. J. M. Kass et al., The global distribution of known and undiscovered ant biodiversity. *Sci. Adv.* **8**, eabp9908 (2022).
62. J. Delabie et al., "Sampling and analysis methods for ant diversity assessment" in *Measuring Arthropod Biodiversity* (Springer, 2021), pp. 13–54.
63. P. Martin, E. Martinez-Ansemil, A. Pinder, T. Timm, M. J. Wetzel, "Global diversity of oligochaetous clitellates ('Oligochaeta'; Clitellata) in freshwater" in *Freshwater Animal Diversity Assessment, Developments in Hydrobiology*, E. V. Balian, C. Lévêque, H. Segers, K. Martens, Eds. (Springer, Netherlands, 2008), pp. 117–127.
64. GBIF Secretariat, "GBIF backbone taxonomy. Checklist dataset" (2021) (21 October 2022).
65. R. M. Schmelz, R. Collado, Checklist of taxa of Enchytraeidae (Oligochaeta): An update. *Soil Org.* **87**, 149–153 (2015).
66. J. T. Lennon, K. J. Locey, More support for Earth's massive microbiome. *Biol. Direct* **15**, 1–6 (2020).
67. F. Rohwer, Global phage diversity. *Cell* **113**, 141 (2003).
68. J. C. Ignacio-Espinoza, S. A. Solonenko, M. B. Sullivan, The global virome: Not as big as we thought? *Curr. Opin. Virol.* **3**, 566–571 (2013).
69. P. Baldrian, T. Větrovský, C. Lepinay, P. Kohout, High-throughput sequencing view on the magnitude of global fungal diversity. *Fungal Diversity* **114**, 539–547 (2022).
70. P. D. Schloss, R. A. Girard, T. Martin, J. Edwards, J. C. Thrash, Status of the archaeal and bacterial census: An update. *mBio* **7**, e00201-16 (2016).
71. W. Foissner, "Protist diversity and distribution: Some basic considerations" in *Protist Diversity and Geographical Distribution, Topics in Biodiversity and Conservation*, W. Foissner, D. L. Hawksworth, Eds. (Springer, Netherlands, 2009), pp. 1–8.
72. X. Li, J. J. Wiens, Estimating global biodiversity: The role of cryptic insect species. *Syst. Biol.* **72**, 391–403 (2022).
73. S. Begall, H. Burda, C. E. Schleich, "Subterranean rodents: News from underground" in *Subterranean Rodents* (Springer, 2007), pp. 3–9.
74. L. Deharveng, C. A. D'Haese, A. Bedos, "Global diversity of springtails (Collembola; Hexapoda) in freshwater" in *Freshwater Animal Diversity Assessment* (Springer, 2007), pp. 329–338.
75. K. T. Ryder Wilkie, A. L. Mertl, J. F. A. Traniello, Species diversity and distribution patterns of the ants of Amazonian Ecuador. *PLoS One* **5**, e13146 (2010).
76. G. M. Barker, *The Biology of Terrestrial Molluscs* (CABI, 2001).
77. C. Lydeard et al., The global decline of nonmarine mollusks. *BioScience* **54**, 321–330 (2004).
78. A. P. Camargo et al., IMG/VR v4: An expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* **51**, D733–D743 (2023).
79. H. Ma et al., The global distribution and environmental drivers of aboveground versus belowground plant biomass. *Nat. Ecol. Evol.* **5**, 1110–1122 (2021).
80. D. E. McAllister, A. L. Hamilton, B. Harvey, E. Don, *Global Freshwater Biodiversity: Striving for the Integrity of Freshwater Ecosystems* (Sea Wind: Bulletin of Ocean Voice International, 1997), **vol. 11**.
81. G. Zotz, P. Weigelt, M. Kessler, H. Kreft, A. Taylor, EpiList 1.0: A global checklist of vascular epiphytes. *Ecology* **102**, e03326 (2021).
82. J.-P. Hugot, P. Baujard, S. Morand, Biodiversity in helminths and nematodes as a field of study: An overview. *Nematology* **3**, 199–208 (2001).
83. F. Mahé et al., Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* **1**, 1–8 (2017).
84. W. Appeltans et al., The magnitude of global marine species diversity. *Curr. Biol.* **22**, 2189–2202 (2012).
85. E. Abebe, W. Decraemer, P. De Ley, "Global diversity of nematodes (Nematoda) in freshwater" in *Freshwater Animal Diversity Assessment, Developments in Hydrobiology*, E. V. Balian, C. Lévêque, H. Segers, K. Martens, Eds. (Springer, Netherlands, 2008), pp. 67–78.
86. W. Xiong et al., A global overview of the trophic structure within microbiomes across ecosystems. *Environ. Intern.* **151**, 106438 (2021).
87. D. Singer et al., Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environ. Intern.* **146**, 106262 (2021).

88. E. Lara, D. Singer, S. Geisen, Discrepancies between prokaryotes and eukaryotes need to be considered in soil DNA-based studies. *Environ. Microbiol.* **24**, 3829–3839 (2022).
89. T. Leho *et al.*, Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
90. M. Öpik *et al.*, The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol.* **188**, 223–241 (2010).
91. E. Egidi *et al.*, A few Ascomycota taxa dominate soil fungal communities worldwide. *Nat. Commun.* **10**, 2369 (2019).
92. C. Averill *et al.*, Defending Earth's terrestrial microbiome. *Nat. Microbiol.* **7**, 1–9 (2022).
93. K. G. Peay, P. G. Kennedy, J. M. Talbot, Dimensions of biodiversity in the Earth mycobiome. *Nat. Rev. Microbiol.* **14**, 434–447 (2016).
94. A. P. Gryganskyi *et al.*, The early terrestrial fungal lineage of conidiobolus–Transition from saprotroph to parasitic lifestyle. *J. Fungi.* **8**, 789 (2022).
95. S. I. Glassman, J. B. H. Martiny, Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *mSphere* **3**, e00148-18 (2018).
96. C. Quast *et al.*, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
97. L. R. Thompson *et al.*, A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
98. P. F. Kemp, J. Y. Aller, Bacterial diversity in aquatic and other environments: What 16S rDNA libraries can tell us. *FEMS Microbiol. Ecol.* **47**, 161–177 (2004).
99. S. Chibani-Chennoufi, A. Bruttin, M.-L. Dillmann, H. Brüssow, Phage-host interaction: An ecological perspective. *J. Bacteriol.* **186**, 3677–3686 (2004).
100. G. F. Hatfull, Mycobacteriophages: Windows into tuberculosis. *PLoS Pathog.* **10**, e1003953 (2014).
101. L. M. Kasman, L. D. Porter, "Bacteriophages" in StatPearls (StatPearls Publishing, 2022) (8 February 2023).
102. K. E. Kortright, B. K. Chan, P. E. Turner, High-throughput discovery of phage receptors using transposon insertion sequencing of bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18670–18679 (2020).
103. I. Hewson *et al.*, Temporal dynamics and decay of putatively allochthonous and autochthonous viral genotypes in contrasting freshwater lakes. *Appl. Environ. Microbiol.* **78**, 6583–6591 (2012).
104. C. Santos-Medellín *et al.*, Spatial turnover of soil viral populations and genotypes overlain by cohesive responses to moisture in grasslands. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2209132119 (2022).
105. A. M. ter Horst *et al.*, Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations. *Microbiome* **9**, 233 (2021).
106. D. M. Durham *et al.*, Substantial differences in soil viral community composition within and among four Northern California habitats. *ISME Commun.* **2**, 1–5 (2022).
107. C. Santos-Medellín, S. J. Blazewicz, J. Pett-Ridge, J. B. Emerson, Viral but not bacterial community succession is characterized by extreme turnover shortly after rewetting dry soils. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.02.12.528215 (Accessed 20 March 2023).
108. W. E. H. Blum, S. Zechmeister-Boltenstern, K. M. Keiblinger, Does soil contribute to the human gut microbiome? *Microorganisms* **7**, 287 (2019).
109. G. Dominguez-Huerta *et al.*, Diversity and ecological footprint of Global Ocean RNA viruses. *Science* **376**, 1202–1208 (2022).
110. A. A. Pratama, J. D. Van Elsas, The 'neglected'soil virome–potential role and impact. *Trends Microbiol.* **26**, 649–662 (2018).
111. S. Roux, J. B. Emerson, Diversity in the soil virosphere: To infinity and beyond? *Trends Microbiol.* **30**, 1025–1035 (2022).
112. C. A. Guerra *et al.*, Blind spots in global soil biodiversity and ecosystem function research. *Nat. Commun.* **11**, 3870 (2020).
113. D. Naylor, R. McClure, J. Jansson, Trends in microbial community composition and function by soil depth. *Microorganisms* **10**, 540 (2022).
114. G. W. Griffith, Do we need a global strategy for microbial conservation? *Trends Ecol. Evol.* **27**, 1–2 (2012).
115. M. A. Anthony, Git repository containing all data and scripts. GitLab. https://gitlab.com/fungalecology/soil_biodiversity_review. Deposited 21 March 2023.