Supporting information

Enumerating soil biodiversity

Mark A. Anthony[1, 2, †], Franz S. Bender[1,3], Marcel G.A. van der Heijden[1,3]

[1]Plant-Soil Interactions Unit, Agroscope, Reckehnholz, Switzerland
[2]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
[3]Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland
[†]Corresponding author (manthony5955@gmail.com)

Table S1. Fraction of protist life so far identified in different ecosystem types across the globe from dataset 1. The fraction of observed protist 18S rDNA amplicon sequence variants separated by ecosystem types. All samples were rarified to 1,000 sequences per sample, and observed species richness is shown. The percent of ecosystem-specific species richness was calculated as the proportion of richness within an ecosystem divided by the total observed species richness. Raw data comes from Xiong et al. (2021) (1).

Ecosystem	**Richness**	**Richness prop.**
Citywater	64	0.112
Lake	68	0.120
Marine	195	0.343
Soils	242	0.425

Table S2. Fraction of protist life so far identified in different ecosystem types across the globe from dataset 2. The fraction of observed protist 18S rDNA operational taxonomic units clustered using SWARM and separated by ecosystem types. All samples were rarified to 10,000 sequences per sample, and observed species richness is shown. The percent of ecosystem-specific species richness was calculated as the proportion of richness within an ecosystem divided by the total observed species richness. Raw data comes from Singer et al. (2021) (2).

Ecosystem	**Richness**	**Richness prop.**
Marine	12540	0.311
Freshwater	11490	0.285
Soil	16337	0.405

Table S3. Fraction of bacterial life so far identified in different ecosystems across the globe from dataset 1. The fraction of observed bacterial 97% 16S rDNA sequence similarity operational taxonomic units (OTUs) was calculated using the SILVA database separated by ecosystem type. The % of total shows the proportion of bacterial OTUs specific to each ecosystem after accounting for variation in the number of sequences. Raw data comes from Schloss et al. (2016) (3).

Ecosystem	**Sequences**	**OTUs**	**OTUs/Seq**	**Prop. of OTUs**
Aerosol	3472	1068	0.308	0.151
Aquatic	214085	43935	0.205	0.101
Built	108799	30012	0.276	0.135
Plant-associated (root)	19695	5052	0.257	0.126
Plant-associated (not root)	14645	4602	0.314	0.154
Soil	74870	23333	0.312	0.153
Host-associated (not plant)	804585	50565	0.063	0.031
Other	19414	5930	0.305	0.150
Total Soil	94565	28385	0.568	0.279

Table S4. Fraction of prokaryotic life so far identified in different ecosystem types across the globe from dataset 2. The fraction of observed bacterial and archaeal 16S rDNA amplicon sequence variants calculated using data from the Earth Microbiome Project separated by ecosystem type. All samples were rarified to 5,000 sequences per sample, and observed species richness is shown. The percent of ecosystem-specific species richness was calculated as the proportion of richness within an ecosystem divided by the total observed species richness. Raw data comes from Thompson et al. (2017) (4).

Ecosystem	**Richness**	**Richness prop.**
Aerosol (non-saline)	272	0.037
Animal corpus	38	0.005
Animal distal gut	333	0.045
Animal proximal gut	173	0.023
Animal secretion	135	0.018
Animal surface	218	0.029
Hypersaline (saline)	282	0.038
Plant corpus	81	0.011
Plant rhizosphere	1397	0.188
Plant surface	173	0.023
Sediment (non-saline)	1178	0.158
Sediment (saline)	880	0.118
Soil (non-saline)	970	0.131
Surface (non-saline)	312	0.042
Surface (saline)	349	0.047
Water (non-saline)	447	0.060
Water (saline)	194	0.026
All soil	2367	0.319

Table S5. Fraction of bacterial life so far identified in different ecosystem types across the globe from dataset 3. The fraction of observed bacterial 'phylotypes' defined as identical groups of PCR products based on 16S rDNA clone libraries in different ecosystem types, including soil. Two non-parametric estimates of species richness are included ACE and Chao1. The *n* columns shows the number of libraries available, and we corrected predictions of species richness using *n* (ACE/*n*, Chao1/*n*). We then calculated the proportion of species richness for ACE and Chao1 in each ecosystem type and computed the average proportion. Raw data comes from Kemp and Aller (2004) (5).

Ecosystem	***n***	***ACE***	***Chao1***	***ACE/n***	***Chao1/n***	***ACE prop.***	***Chao1 prop.***	***Average prop.***
Bacterioplankton	51	69	50	1.353	0.980	0.015	0.014	0.014
Biofilms	30	60	44	2.000	1.467	0.022	0.020	0.021
Gas hydrates	6	91	42	15.167	7.000	0.165	0.098	0.131
Groundwater	17	114	81	6.706	4.765	0.073	0.067	0.070
Hyperthermal	18	82	38	4.556	2.111	0.050	0.029	0.040
Sediment	30	107	70	3.567	2.333	0.039	0.033	0.036
Suspended particles	8	51	33	6.375	4.125	0.069	0.058	0.064
Digestive systems	18	104	129	5.778	7.167	0.063	0.100	0.082
Compost	6	164	124	27.333	20.667	0.298	0.289	0.293
Soils	19	293	376	15.421	19.789	0.168	0.276	0.222
Bioreactors	22	79	26	3.591	1.182	0.039	0.017	0.028

Table S6. Fraction of archaeal life so far identified in different ecosystems across the globe in dataset 1. The fraction of observed archaeal 97% 16S rDNA sequence similarity operational taxonomic units (OTUs) was calculated using the SILVA database separated by ecosystem type. The % of total shows the proportion of archaeal OTUs specific to each ecosystem after accounting for variation in the number of sequences. Raw data comes from Schloss et al. (2016) (3).

Ecosystem	**Sequences**	**OTUs**	**OTUs/Seq**	**Prop of OTUs**
Aquatic	34400	4838	0.141	0.087
Built	7286	1219	0.167	0.104
Plant-associated (root)	200	61	0.305	0.189
Plant-associated (not root)	22	7	0.318	0.198
Soil	7906	1125	0.142	0.088
Host-associated (not plant)	1090	348	0.319	0.198
Other	2565	559	0.218	0.135
Total Soil	8106	1186	0.447	0.278

Table S7. Fraction of archaeal life so far identified in different ecosystem types across the globe from dataset 2. The fraction of observed archaeal 'phylotypes' defined as identical groups of PCR products based on 16S rDNA clone libraries in different ecosystem types, including soil. Species richness was estimated after subsampling to a library size of 120 clones for each ecosystem type. We then calculated the proportion of observed richness in each ecosystem type. Raw data is from Auguet et al. (2010) (6).

Ecosystem	**Richness**	**Richness prop.**
soil	64	0.103
marine plankton	77	0.124
freshwater sediment	81	0.130
marine sediment	90	0.145
hypersaline planktonic environments	97	0.156
hydrothermal vents	101	0.163
freshwater plankton	111	0.179

Table S8. Fraction of viral life so far identified in different ecosystems across the globe. The fraction of observed metagenomic viral contigs (mVC) identified across different ecosystem types after normalizing for variation in sequencing depth (megabases per sample: Mb). The % of mVC shows the proportion of mVCs specific to each ecosystem after accounting for variation in the number sequences. Raw data was compiled from sources in Paez-Espino et al. (2016) (7).

Ecosystem	Megabases (Mb)	mVC	Samples	mVC/Mb	% of mVC
Engineered	10486	7970	154.000	0.760	8.701
Freshwater	96314	20576	178.000	0.214	2.446
Host-associated (human)	10349	30849	664.000	2.981	34.123
Host-associated (other)	23452	5596	99.000	0.239	2.731
Host-associated (plants)	3909	1233	91.000	0.315	3.611
Marine	29602	50587	397.000	1.709	19.562
Non-marine Saline and Alkaline	2825	4538	41.000	1.606	18.388
Terrestrial (other)	1659	343	32.000	0.207	2.367
Terrestrial (soil)	5794	2840	64.000	0.490	5.611
Thermal springs	1828	393	49.000	0.215	2.461

Table S9. List of non-soil inhabiting Enchytraeid species. The species listed here were identified from the genera outlined in van van Vliet (1999)(8).

Grania acanthochaeta Rota & Erséus, 1996
Grania algida Rota & Erséus, 1996
Grania alliata Coates & Stacey, 1993
Grania americana Kennedy, 1966
Grania angustinasus Rota & Erséus, 1996
Grania antarctica Rota & Erséus, 1996
Grania aquitana Rota & Erséus, 2003
Grania ascophora Coates, 1990
Grania atlantica Coates & Erséus, 1985
Grania bekkouchei Prantoni, De Wit & Erséus, 2016
Grania bermudensis Erséus & Lasserre, 1976
Grania brasiliensis Prantoni, De Wit & Erséus, 2016
Grania breviductus De Wit, Rota & Erséus, 2009
Grania bykane Coates, 1990
Grania canaria Rota & Erséus, 2003
Grania capensis Prantoni, De Wit & Erséus, 2016
Grania carchinii Rota & Erséus, 1996
Grania carolinensis Prantoni, De Wit & Erséus, 2016
Grania chilensis Prantoni, De Wit & Erséus, 2016
Grania cinctura De Wit & Erséus, 2007
Grania colorata De Wit, Rota & Erséus, 2009
Grania conjuncta Coates & Stacey, 1993
Grania crassiducta Coates, 1990
Grania cryptica Prantoni, De Wit & Erséus, 2016
Grania curta De Wit & Erséus, 2007
Grania darwinensis (Coates & Stacey, 1997)
Grania dolichura Rota & Erséus, 2000
Grania ersei Coates, 1990
Grania eurystila Coates & Stacey, 1997
Grania fiscellata De Wit & Erséus, 2007
Grania fortunata Rota & Erséus, 2003
Grania fustata De Wit & Erséus, 2007
Grania galbina De Wit & Erséus, 2007
Grania hastula Coates, 1990
Grania hinojosai Prantoni, De Wit & Erséus, 2016
Grania hirsuticauda Rota & Erséus, 1996
Grania homochaeta De Wit, Rota & Erséus, 2009
Grania hongkongensis Erséus, 1990
Grania hylae Locke & Coates, 1999
Grania hyperoadenia Coates, 1990
Grania incerta Coates & Erséus, 1980
Grania inermis Erséus, 1990
Grania integra Coates & Stacey, 1997
Grania lasserrei Rota & Erséus, 1997
Grania la×artus Locke & Coates, 1999
Grania levis Coates & Erséus, 1985
Grania longiducta Erséus & Lasserre, 1976
Grania longistyla Coates & Stacey, 1993
Grania macrochaeta (Pierantoni, 1901)
Grania maricola Southern, 1913
Grania mauretanica Rota & Erséus, 2003
Grania mira Locke & Coates, 1998
Grania monochaeta (Michaelsen, 1888)
Grania monospermatheca Erséus & Lasserre, 1976
Grania novacaledonia De Wit & Erséus, 2007
Grania ocarina Rota, Erséus & Wang, 2003
Grania occulta De Wit & Erséus, 2010
Grania ovitheca Erséus, 1977
Grania pacifica Shurova, 1979
Grania papillata De Wit & Erséus, 2007
Grania papillinasus Rota & Erséus, 2003
Grania parvitheca Erséus, 1980
Grania paucispina (Eisen, 1904)
Grania postclitellochaeta (Knöllner, 1935)
Grania principissae (Michaelsen, 1907)
Grania pusilla Erséus, 1974
Grania quaerens Rota, Wang & Erséus, 2007
Grania reducta Coates & Erséus, 1985
Grania regina De Wit, Rota & Erséus, 2009
Grania roscoffensis Lasserre, 1967

Grania simonae Prantoni, De Wit & Erséus, 2016
Grania sperantia Rota, Wang & Erséus, 2007
Grania stephensoniana Rota & Erséus, 1997
Grania stilifera Erséus, 1990
Grania tasmaniae Rota & Erséus, 2000
Grania torosa Rota & Erséus, 2003
Grania trichaeta Jamieson, 1977
Grania unitheca Prantoni, De Wit & Erséus, 2016
Grania vacivasa Coates & Stacey, 1993
Grania variochaeta Erséus & Lasserre, 1976
Grania vikinga Rota & Erséus, 2003
Aspidodrilus kesalli
Aspidodrilus eburneensis
Barbidrilus paucisetus
Randidrilus Coates & Erséus, 1985
Randidrilus codensis (Lasserre, 1971)
Randidrilus quadrithecatus Coates & Erséus, 1985
Randidrilus westheidei (Kossmagk-Stephan, 1983)
Stephensoniella Cernosvitov, 1934
Stephensoniella Lastochkin, 1935 accepted as Potamodrilus Lastochkin, 1935
Stephensoniella fluviatilis Lastochkin, 1935 accepted as Potamodrilus fluviatilis (Lastochkin, 1935)
Stephensoniella marina (Moore, 1902)
Stephensoniella sterreri (Lasserre & Erséus, 1976)
Stephensoniella trevori (Coates, 1980)
Pelmatodrilus
Propappus glandulosus Michaelsen, 1905
Propappus volki Michaelsen, 1916
Propappus arhyncotus Sokolskaja, 1972

Table S10. Footnotes of methodological details for the estimations of biodiversity shown outside of the parentheses using letters in Table 1.

[a] Bagyaraj et al. (2016) (9) claims 50% of nematodes live in soil. Seven percent of the currently described 2.7×10^4 Nematoda species live in fresh-water(10), and 50% live in marine environments (9). If this 57% reflects the non-soil inhabiting species, then we can expect 43% of Nematoda biodiversity to be found in soil (or inside animals and plants that live in soil), an estimate not far off from the expert opinion of Bagyaraj et al. (2016) (9). We therefore use 43% for the lower prediction, 50% for the upper prediction, and the average of the two for the central prediction 46.5%. We also note that this value could be an under-estimate since so few studies have used appropriate molecular tools to characterize nematode biodiversity at a global scale.

[b] This value was calculated using data from tropical tree canopies. This is expected to capture most Arthropoda life since tropical species make the largest contribution to global species richness. There is a ratio of 162 host-specific beetles per tree species, which suggests that there may be as many as 3×10^7 tropical arthropods (11). Yet, this value has been the subject of debate as an over-estimation. Others suggests that only 13.5% of beetle species are actually host specific (12), and they use this percentage to extrapolate total species richness based on the total number of rainforest tree species, which is potentially more accurate (13). Indeed, some report that an estimate of 3×10^7 species has a <0.00001 probability of being true(14). Further, some have suggested that the diversity in soil may be an order of magnitude higher than that of tree canopies, ultimately providing a solid basis to be critical of Erwin (1982)'s (12) estimation. So far, few studies assessed the richness of arthropods in the soil and this number could be higher. Note that the upper Insecta value is higher than the total Arthropoda estimate even though insects are a group of arthropods. This can occur when two highly uncertain and speculative estimates are provided, and we are unable to differentiate which may be more accurate than the other, so we retain both. We emphasize this uncertainty using questions marks (see below).

[c] These are estimates derived from the proportion of life in soil that has already been predicted at the lower end. The lower end prediction was used to calculate the central and upper predictions of soil life using the global species prediction. Where there was a lower and central prediction already for the soil, the mean proportion of soil life for both was used to calculate the upper estimate.

[d] Bagyaraj et al. (2016) (9) claims that there are 50,000 mites (a group of Arachnida) species known to live in soil based on expert opinion. This value does not include all Arachnida inhabiting soil. Nevertheless, this value is similar to the lower value estimate proposed by Decaëns et al. (2006) (15), and it is the best estimate currently available.

[e] From current knowledge of aquatic collembola inventories and extrapolation techniques, there are 525 species estimated (414 freshwater, 109 marine, and 2 marine/freshwater species (16). This value can therefore be subtracted to derive the value from the lower and upper estimates of total numbers to generate soil-dwelling predictions.

[f] All termites but the Kalotermitidae and Termopsinae live in association with soil (17). Globally, there are 450 Kalotermitidae and 20 Termopsidae. Thus, we can subtract 470 from the global species estimates to derive the soil-inhabiting species.

[g] Ryder et al. (2010) (18) found that ca. 50% of ant species live in or on soil in a western Amazonian rainforest of Ecuador. This is a biodiversity hotspot in the tropics and may generally represents above versus belowground ant habitat stratification. To ascertain the low- and high-end estimates of ant richness, we can assume 50% of the global estimate.

[h] A compilation of species was identified in the literature (see SI Appendix, Table S9) for all the soil inhabiting genera listed in van Vliet (2000) (8). We identified a total of 98 species, which we can then subtracted from the total species list to derive estimates of soil-dwelling species.

[i] This estimate is for total terrestrial biodiversity and includes species that live on or within soil litter and other terrestrial ecosystems, though sampling of mollusks is typically done by hand or after collecting micro-mollusks in litter and soil passed through a 5 mm sieve (e.g. (19)), suggesting a primarily soil origin.

[j] Current estimates suggest that 7.2-11.4% of angiosperm families are entirely or partially aquatic(20). There are also approximately 31,311 vascular epiphytes in the world, 90% of which are angiosperms (21). Assuming that 9.3% of angiosperm species are aquatic, this can be subtracted from the angiosperm component of total plant biodiversity (79% of known plant species are angiosperms (22)). Seventy-nine percent of total plant diversity (low end total = 380,000) = 300,200 (angiosperms); 9.3% of angiosperms = 27,920 (aquatic angiosperms); aquatic angiosperms (27,900) + epiphytes (31,311) = 59,211 non-soil inhabiting plant species. Subtracting this value allows prediction of soil-inhabiting plant species. Assuming there are 537,000 plant species gives us (0.79 * 537,000 = 424,230 angiosperms; 0.093 * 424,230 = 39,453; 537,000 – (39,453 + 31,311) = 466,236). Since these plants have 22-67% of their biomass in soil(11) and have seeds or spores that germinate in soil, we can consider them soil inhabiting.

[k] Microbe is an ambiguously used term in the literature. Here, it is used to represent studies that include bacteria, archaea, and fungi, but not studies that strictly consider bacteria. In the lower and central estimates, this is the sum of bacteria, archaea, and fungi estimates. If one of the bacterial, archaeal, or fungal predictions are not available, this total is not calculated (i.e. upper, soil microbial diversity). This number should also include protists (23) but because the studies included here to estimate the upper range of microbial biodiversity did not include protists, we also did not add protists in order to make comparisons across the lower, central, and upper predictions equivalent. Note that one can simply add the protist value to the total microbe value in order to estimate total microbe richness including protists.

[l] There are a few hundred to a few thousand viral protein clusters in the pan-genome of individual bacterial species (24) and hundred to thousands of unique viral operational taxonomic units (vOTUs) found per bacterial species (Figure S1A). Assuming most viruses do not infect multiple hosts or share genes across host species (Figure S1A), one can multiply the total number of microbes by the number of viral taxa found per species (i.e. here we choose an intermediate value of 1,000) to estimate an upper prediction of phage diversity. Using the upper estimate of bacterial diversity (3.7×10^9) and a ratio of 1000:1, we predicted the upper and lower ranges of viral diversity.

[m] Based on the distribution of vOTU richness across ecosystem types to-date (Figure S1B), we found that 9.9% of all vOTUs were from soil (Figure S1C). We arrived at a slightly smaller estimate of 5.6% using an independent dataset (see main text; SI Appendix, Table S8). We therefore estimate that a minimum of 6-10% of viral diversity is found in soil, though this number is likely to be much higher. Because the proportion of bacterial life in soil is ca. 43%, phage should theoretically track this estimate. We therefore use a 6% lower, 9.9% central, and 43% upper estimate (our central estimate for the proportion of bacterial life in soil, used here because we assume most soil viruses are phage) to compute the proportion of viral life in soil relative to global estimates.

[n] Wien (2021) (25) re-analyzed data from Larsen et al. (2017) (26) using updated concepts of host specificity, and it shows that the number of unique bacterial species found per insect species is in the 1,000's range. They then use this number to re-estimate global species predictions for all groups, but highlighting bacteria.

[o] We analyzed three global prokaryotic datasets to estimate the proportion of observed bacterial species in soil versus other ecosystem types. We first re-analyzed data from the SILVA database(27) with meta-data organized by others (3) to estimate that 27.9% of all 97% OTUs have been observed in soil, after correcting for sequencing effort (see SI Appendix, Table S3). We then repeated this process using data from the Earth Microbiome Project (4) and found 32.9% of amplicon sequence variants, 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 (5) to find 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(28). Curtis et al. (2002) (28) predicts that soil diversity would be 4×10^6 while total diversity would be 4.5×10^6. This is equivalent to soils containing 88.8% of the bacterial diversity. An estimated total of 4.5×10^6 is also not very different from the 4.3×10^6 estimate of Louca et al. (2019)(29). Estimates of the proportion of bacterial life in soil therefore range between 22-89% with an average (hereafter: central estimate) of 43%. We therefore used 22.2, 43, and 88.8% (average = 51%, as shown in Figure 2) as our low, central, and upper predictions of the proportion of bacterial species in soil.

[p] This value is based on extrapolation of fungal species observations in typical insect hosts generalized to include all animals. It does not thoroughly account for overlap among taxa across animal hosts nor does it include other hosts of fungi, such as plants. It also does not account for free-living fungi.

[q] Using values from Schmit and Mueller (2007) (30) there are estimated 3,000 lichenicolous fungi, 8,400 aquatic fungi, 20,000 arthropod associated fungi, 1,200 microsporidians (mostly animal parasites) which equals 32,600 non-soil dwelling fungal taxa. We estimated that animal-associated fungal diversity is 10× lower than soil fungal diversity using data presented in Peay et al. (2016) (31). Thus, if we account for the 32,600 non-soil dwelling fungi, we can assume that 10% of remaining fungal diversity is animal-associated. If these fungi lack dual-capacities to live in animal hosts and the soil environment, an assumption that we know is not entirely correct, then there are 2.2×10^5 and 6.2×10^5 species that do not live in soil at the low and central estimates, respectively. It is worth noting that soil biodiversity would be even higher if animal-associated species were included as facultative saprotrophs in soil. Subtracting the arthropod-associated and microsporidians (21,200) further constrains this value to 198,800 and 598,800 non-soil inhabiting fungi at the lower and central estimates, respectively. Subtracting total fungal diversity from these estimates derives soil-inhabiting fungal biodiversity. Note that this derivation was not applied for the upper prediction because this estimate is already based on arthropod-associated predictions of fungal biodiversity. It has received considerable scrutiny, but it is the highest predicted estimate to date, theoretically possible, and therefore included as an upper limit in our study.

[r] Louca et al. (2019) (29) identified a range of 7 to 14×10^4 archaeal OTUs based on the V4 region using the tWLRM and breakaway estimation methods, respectively. They also find that there are 2.7 times more OTUs discovered when analyzing the full-length 16S region relative to the V4 region. Thus, estimates were multiplied by 2.7 to reach the final value.

[s] We analyzed two independent datasets to estimate the proportion of observed archaeal species in soil versus other ecosystem types. We show that the proportion of Archaea OTUs in soil to range from 10.3-27.8% with average of 19.1% (Tables S6-7). We therefore use 10.3, 19.1, and 27.8% as our lower, central, and upper soil proportion estimates.

[t] 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 40.4% and 42.5% of the protist taxa were found in soil compared to aquatic and marine ecosystem types in our two datasets (1, 2). We therefore used the low percent in soil (40.4) and the average of two (41.45) to compute the fraction of protist species found in soil relative to global predictions. Note that we did not do this for the upper prediction of protist diversity in soil because this estimate is based on arthropod-associated predictions of protist biodiversity, and it has received considerable scrutiny. It is the highest predicted estimate to date, theoretically possible, and therefore included as an upper limit in our study.

[?] Denotes that this estimate is highly speculative and controversial in the literature. Note that this is different from normal-sized question marks.

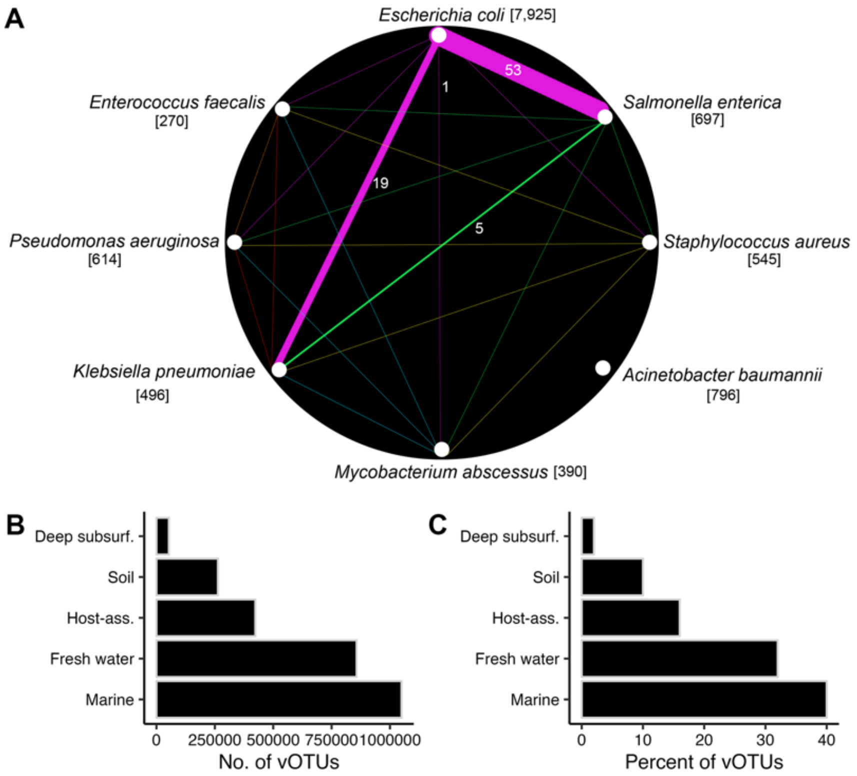

Figure S1. Overview of viral operational taxonomic unit (vOTU)s identified and shared within bacterial hosts and across ecosystem types. (**A**) Co-occurrences of shared vOTUs among bacterial host species with >1,000 viral metagenomes in the IMG/VR database (v4.1, IMG_VR_2022-09-20_7.1). Values along lines and in square brackets show the number of vOTUs shared among bacterial hosts and the total number of vOTUs thus identified within each host species, respectively. (**B**) The number of unique vOTUs identified to-date across deep subsurface, soil, host-associated, fresh water, and marine ecosystems, and (**C**) the percent of vOTUs identified to-date in each ecosystem type. Abbreviations: Deep subsurf. = deep subsurface; Host-ass. = host-associated

References

1. W. Xiong, *et al.*, A global overview of the trophic structure within microbiomes across ecosystems. *Environment International* **151**, 106438 (2021).

2. D. Singer, *et al.*, Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environment International* **146**, 106262 (2021).

3. 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).

4. L. R. Thompson, *et al.*, A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).

5. P. F. Kemp, J. Y. Aller, Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiology Ecology* **47**, 161–177 (2004).

6. J.-C. Auguet, A. Barberan, E. O. Casamayor, Global ecological patterns in uncultured Archaea. *ISME J* **4**, 182–190 (2010).

7. D. Paez-Espino, *et al.*, Uncovering Earth's virome. *Nature* **536**, 425–430 (2016).

8. P. C. J. Van Vliet, Enchytraeids. *Handbook of Soil Science* **58**, C-70–76 (1999).

9. 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.

10. 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.

11. H. Ma, *et al.*, The global distribution and environmental drivers of aboveground versus belowground plant biomass. *Nat Ecol Evol* **5**, 1110–1122 (2021).

12. T. L. Erwin, Tropical forests: their richness in Coleoptera and other arthropod species. *The Coleopterists Bulletin* (1982).

13. G. H. R. Osler, A. J. Beattie, Contribution of oribatid and mesostigmatid soil mites in ecologically based estimates of global species richness. *Austral Ecology* **26**, 70–79 (2001).

14. A. J. Hamilton, *et al.*, Quantifying Uncertainty in Estimation of Tropical Arthropod Species Richness. *The American Naturalist* **176**, 90–95 (2010).

15. T. Decaëns, J. J. Jiménez, C. Gioia, G. Measey, P. Lavelle, The values of soil animals for conservation biology. *European Journal of Soil Biology* **42**, S23–S38 (2006).

16. 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.

17. T. G. Wood, Termites and the soil environment. *Biol Fert Soils* **6**, 228–236 (1988).

18. 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).

19. T. Waki, Diversity of terrestrial mollusks and their helminths in artificial environments in Yoyogi Park, Tokyo, Japan. *Journal of Asia-Pacific Biodiversity* **10**, 254–256 (2017).

20. 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* **11** (1997).

21. G. Zotz, P. Weigelt, M. Kessler, H. Kreft, A. Taylor, EpiList 1.0: a global checklist of vascular epiphytes. *Ecology* **102**, e03326 (2021).

22. M. J. M. Christenhusz, J. W. Byng, The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201–217 (2016).

23. D. A. Caron, A. Z. Worden, P. D. Countway, E. Demir, K. B. Heidelberg, Protists are microbes too: a perspective. *ISME J* **3**, 4–12 (2009).

24. J. C. Ignacio-Espinoza, S. A. Solonenko, M. B. Sullivan, The global virome: not as big as we thought? *Current opinion in virology* **3**, 566–571 (2013).

25. J. J. Wiens, Vast (but avoidable) underestimation of global biodiversity. *PLOS Biology* **19**, e3001192 (2021).

26. 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. *The Quarterly Review of Biology* **92**, 229–265 (2017).

27. C. Quast, *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**, D590–D596 (2012).

28. T. P. Curtis, W. T. Sloan, J. W. Scannell, Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences* **99**, 10494–10499 (2002).

29. S. Louca, F. Mazel, M. Doebeli, L. W. Parfrey, A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS biology* **17**, e3000106 (2019).

30. J. P. Schmit, G. M. Mueller, An estimate of the lower limit of global fungal diversity. *Biodiversity and Conservation* **16**, 99–111 (2007).

31. K. G. Peay, P. G. Kennedy, J. M. Talbot, Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology* **14**, 434–447 (2016).