The biomass distribution on Earth

Y. M. Bar-On, R. Phillips, R. Milo

Supplementary Information Appendix

Contents

www.pnas.org/cgi/doi/10.1073/pnas.1711842115

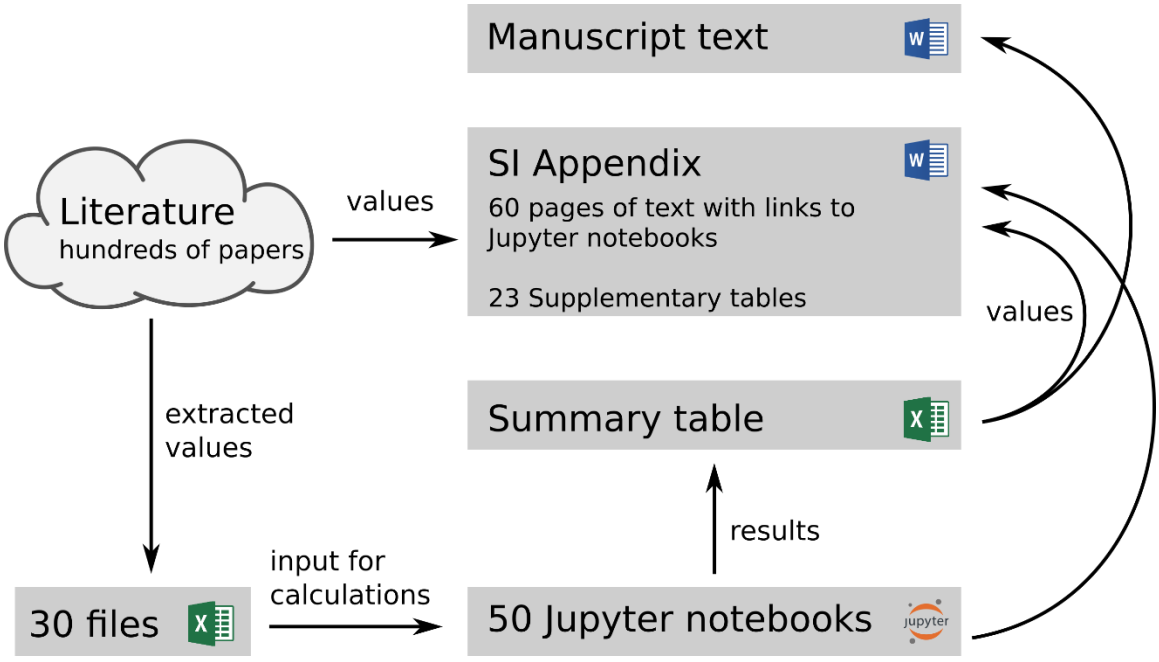
Supplementary Figures and Tables

Taxon-specific detailed description of data sources and procedures for estimating biomass

Overview

Here we provide a detailed description of the data and procedures for arriving at the final estimates presented in the paper. The description is divided into the different groups of organisms. For some organisms, there is further division into the different environments they reside in. We used **bold font** to highlight the concluding values derived for each taxonomic group. All of the data used to generate our estimates, along with the code for analyzing the data, are open-source and available at https://github.com/milo-lab/biomass_distribution.

The *SI Appendix* and all the accompanying files are made accessible so that researchers can see in a completely transparent manner how each value was derived from the many literature sources and be able to update the analysis using extra data or a different data analysis approach. We rely on hundreds of studies from the literature to support the data presented in the *SI Appendix*. To generate our estimates of biomass, we extract values from the literature into spreadsheet files. Our analysis pipeline is comprised of about 50 different Jupyter notebooks which use the data extracted from the literature as input and generate our estimates. The result of our analysis is summarized in a summary table located at https://github.com/milo-lab/biomass_distribution. We use the results of our analysis when reporting the values in the manuscript and *SI Appendix*, including all the associated figures and tables.



Plants

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Biomass

There are several estimates for the biomass of specific components of plant biomass, mainly forests, and they are based on both direct field observations (termed inventory data; 1, 2) and on remote-sensing data (3–5). These studies, however, have mainly focused on forests, and have not quantified the contribution of other components of the global plant biomass, such as shrubs, grasses, or locations with trees not defined as forests (because of an average tree cover of less than 10%). Alternatively, well-established studies (6–8) provide estimates for the carbon density and extent in plants for all biomes, which takes into account also biomes that are not dominated by trees (such as grasslands, shrublands, and savannas). These studies, however, provide estimates for the biomass density of plants in undisturbed ecosystems, and do not take into account the very broad land-use changes which have affected many natural ecosystems. These changes include transformation of natural ecosystems to cropland, as well as the impact of forestry and grazing on natural ecosystems. This is one of the reasons that the global estimates for forest biomass in Saugier et al. or Ajtay et al. diverge from more recent inventory-based or remote-sensing-based estimates (1–5). The best resource we were able to find as a basis for our estimate of the global biomass of plants is Erb et al. (9). Their study assesses the biomass of both forest and non-forest ecosystems and takes into account impacts of land-use changes. We briefly describe the methodology Erb et al. used for estimating the global biomass of plants, for full details we refer the reader to the original paper (9). Erb et al. generated seven maps of plant biomass stocks. Their best estimate for the global biomass of plants is based on the mean of those seven different maps. The first six maps could be divided to two broad categories: two inventory-based maps, and four remote sensing-based maps. We now describe those estimates in more detail.

For the inventory-based maps, Erb et al. consider five main land cover categories: forests, cropland, artificial grassland (forest area converted to pastures or meadows), other wooded land (5-10% tree cover, including savannas and shrubland) and natural grasslands (<5% tree cover). To each land-use unit, Erb et al. assign typical biomass stock density values from the literature or census statistics. In the first map, Erb et al. use national-level data from the global Forest Resource Assessment for the forest land-cover. The second map uses data on forest inventories from Pan et al. (1). Erb et al. set the biomass density of grassland-tree mosaics (other wooded land and natural grasslands which contain trees), at 50% of the biomass density of the neighboring forests (forests which are in the same country as the grassland-tree mosaics). Erb et al. based this assumption on data from the FRA on the ratio between other wooded land biomass and forest biomass in ≈70 countries (10). For herbaceous vegetation units (artificial grassland, cropland and natural grassland without trees), Erb et al. assumed biomass stocks to equal the annual amount of net primary production. For permanent cropland, Erb et al. added 3 kg C m^{-2} for tree-bearing systems and 1.5 kg C m^{-2} for shrub bearing systems to account for woody above- and belowground compartments.

For the remote sensing-based maps, Erb et al. combined independent remote-sensing products for tree vegetation and expanded them to account for below-ground and herbaceous compartments where necessary. The main data sources used for the construction of the maps are Thurner et al. (5) for the northern boreal and temperate forests, and Baccini et al. and Saatchi et al. for tropical forests (3, 4). For some regions (the southernmost part of Australia, and parts of Oceania), no remote-sensing data are available. In these regions, Erb et al. used the inventory-based map to fill missing values. The first two remote-sensing maps use either Saatchi et al. or Baccini et al. for the tropical forest component of the map. The additional two remote sensing-based maps use the minimal or maximal biomass densities at each grid cell of the map, respectively. The last map Erb et al. use is the map constructed by Ruesch & Gibbs based on the IPCC tier-1 globally consistent default biomass values (8). To estimate the total biomass of plants from those seven different estimates, we use the best estimate reported by Erb et al., which is ≈450 Gt C (link to full calculation).

The analysis of Erb et al. does not include halophytic flowering plants (plants living in water with high salinity) and bryophytes (liverworts, hornworts and mosses). Biomass of halophytic flowering plants is dominated by mangroves and seagrass. An estimate for the total biomass of mangroves is roughly ≈4 Gt C (11). This value is consistent with estimates for the global storage of organic carbon in mangroves, and the fraction of this carbon that is living mangroves (and not dead biomass or soil organic carbon; 12). For seagrass, Fourqurean et al. (13), estimate a global biomass of ≈0.1 Gt C. As for bryophytes, there seems to be no global data on bryophytes alone, but Elbert et al. (14) give values of ≈5 Gt C for 'cryptogamic covers', i.e. varying mixtures of bryophytes, lichens, eukaryotic algae, cyanobacteria and fungi growing as epiphytes, as crusts on arid lands, and in bogs. This translates into an organic C content of about 1% of the total for terrestrial plants. Supporting these general values, Edwards et al. (15) analyzed the global carbon content of 'cryptogamic covers' in the Ordovician-Silurian period (about 400-500 million years ago) and suggested it was similar to today's values, i.e. a total of about ≈5 Gt C. For the sum of marine macroalgae (green and brown algae), De Vooys (16) gives a total global biomass of of 0.0075 Gt C, based on annual productivity and assuming one turnover of the standing crop each year. Cherpy-Roubaud & Sournia (17) cite global annual productivity of all taxa of marine macroalgae of 2.55 Gt C yr^{-1}, or a global standing crop of 2.55 Gt C assuming one turnover of the standing crop each year (18). This 2-3 orders of magnitude range of values (0.0075-2.55 Gt C) means that more work is needed to obtain tightly-constrained estimates.

Plant biomass is dominated by land plants (embryophytes), and more specifically by vascular plants (tracheophytes), with only a minor contribution from bryophyte biomass (14). Overall, we estimate that the global plant biomass, including contributions from land plants as well as other groups such as bryophytes and all marine plant contributions is **≈450 Gt C**.

We now analyze the associated uncertainty of the estimate for the total biomass of plants. We report the uncertainty as a fold change factor from the mean, representing a range akin to a 95% confidence interval of the estimate. One approach to assess the uncertainty associated with the estimate of the total biomass of plants is to calculate the 95% confidence interval around the

geometric mean of the seven estimates from the seven different maps generated by Erb et al. (9). This yields a rather small uncertainty of ≈1.1-fold (link to full calculation). However, this procedure only includes uncertainty stemming from the variation between different estimates and does not include the systematic uncertainty stemming from assumptions made to produce each one of the seven estimates. The main type of uncertainty we believe is not accounted for by this procedure is the uncertainty in the biomass contribution from other wooded land (such as savannas). In the inventory-based estimates, the biomass density of other wooded land is assumed to be ≈50% of the national average biomass density of forests. This assumption has uncertainty associated with it which is not easily quantified. Similarly, the remote sensing-based estimates are mainly designed to quantify the biomass of forests, and therefore there is significant uncertainty regarding the measurement of non-forest tree and shrub plant forms by remote sensing. Providing a rigorous quantification of this type of uncertainty is hard. To account, even if partially, for this uncertainty, we use the multiplicative ratio between the upper (and lower) most estimate relative to our best estimate, which is ≈1.2-fold, as our best projection of the uncertainty associated with our estimate of the biomass of plants (link to full calculation).

Biomass of roots and leaves

Plant tissues are composed of an extracellular scaffold made out of cell wall (mainly cellulose and lignin), supporting a network of cytoplasmic space, termed the protoplasm. Conceptually, plants are not different from other organisms, which also contain supporting tissues such as endo- or exoskeletons, or even the extracellular matrix of microbial biofilms. The ratio between protoplasm and cell wall varies between plant compartments, with leaves containing the least amount of supporting tissues and stems of woody plants (such as trees) mainly composed of supporting tissues. To estimate the total biomass of non-woody tissues, we chose to remove the biomass of stems tissue, as it is dominated by cell wall. Roots also have a non-negligible fraction of cell wall biomass, but the ratio between cell wall and protoplasm in roots is harder to estimate globally. Therefore, to calculate the non-woody plant biomass fraction, we consider only the biomass of leaves and roots. In order to estimate the total biomass of roots and leaves, we rely on two independent methods. The first relies on a meta-analysis of the biomass allocation to different plant compartments across biomes (19). Using these data, we calculated the average biomass fraction of leaves and roots out of the total biomass to be ≈30% by taking into account the contribution of each biome to the global plant biomass (20; link to full calculation). This means that out of the 450 Gt C of plant biomass, 30%, or ≈150 Gt C is concentrated in metabolically active plant tissues.

Our second method for estimating the non-woody plant biomass combines estimates for the global biomass of leaves and roots. For the global root biomass, we rely on the estimate from Jackson et al. (21), of ≈150 Gt C. For the global leaf biomass, we calculate the total leaf area of forests, which dominate plant biomass, by using a combination of biome level estimates for the leaf area index (LAI; 22) and the total forest area in each biome (23). We then convert the total area of leaves to an estimate of the total dry mass of leaves by using the Glopnet database (24), which measured leaf mass per area (LMA) for ≈2500 plant species. This independent procedure yields an estimate of ≈30 Gt dry weight, or ≈15 Gt C assuming 50% carbon content (link to full calculation). Summing our estimate for the total biomass of roots and leaves, we get ≈160 Gt C.

We use the geometric mean of both methods, which is ≈150 Gt C, as our best estimate of the total non-woody plant biomass.

To estimate the total belowground plant biomass, we rely on the same two methods as for the total non-woody biomass. We now consider only root biomass. The first method yields an estimate of ≈120 Gt C (link to full calculation). Our second source is the estimate of the global root biomass from Jackson et al. (21). We use the geometric mean of both estimates, which is ≈130 Gt C, as our best estimate for the belowground plant biomass. This puts the aboveground plant biomass at ≈320 Gt C.

Comparison of plant and bacterial biomass

Our best estimates for the total biomass of plants and bacteria suggests that these taxa account for ≈80% and ≈10% of the global biomass of the biosphere, respectively. The uncertainty associated with our estimate of the biomass of plants is relatively small (≈1.2-fold), and the uncertainty associated with our estimate of the biomass of bacteria is much larger (≈9-fold). In order to quantify the certainty in the statement that plants are more dominant in terms of biomass than bacteria, we use a bootstrapping approach. We randomly sample from the distribution of our estimates for the total biomass of plants and bacteria (with the width of the distribution corresponding to the level of our uncertainty). We find that in ≈90% of cases plant biomass is higher than the biomass of bacteria (link to full calculation).

Bacteria and Archaea (Prokaryotes)

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

One of the best-known estimates of global biomass of prokaryotes (bacteria and archaea) is the meta-analysis study by Whitman et al. (25), which analyzed data of cell abundance densities for various environments and extrapolated the total prokaryotic biomass using estimates of cell mass in each environment. We used this study as a baseline estimate on which various corrections and updates were imposed for the different environments, as highlighted below. An estimate for the distribution of biomass between bacteria and archaea is not available currently. We generate estimates for the biomass of bacteria and archaea in a two-step process. First, we estimate the total biomass of prokaryotes in each environment, and then we estimate the fraction of bacteria and archaea out of the total biomass of this environment. Combining the contributions from each environment, we estimate that the global biomass of bacteria is ≈70 Gt C, which is dominated by ≈60 Gt C of terrestrial deep subsurface bacteria. We estimate the global biomass of archaea at ≈7 Gt C, with ≈4 Gt C and ≈3 Gt C contributed by the terrestrial and marine deep subsurface, respectively. In addition to estimating the biomass of prokaryotes in each environment, we also present in detail the uncertainties associated with each estimate. When combining the uncertainties for the biomass of bacteria and archaea in different environments we estimate the uncertainty of the global biomass of bacteria and archaea to be about 10-fold and 13-fold, respectively, dominated by the uncertainty of the biomass of terrestrial deep subsurface bacteria and archaea.

Marine

To generate an estimate of the biomass of marine bacteria and archaea (prokaryotes), we first estimate the total number of marine prokaryotes, and the characteristic carbon content of a single marine prokaryote. We generate our estimate for the biomass of marine prokaryotes by multiplying the total number of cells by the characteristic carbon content per cell. Our estimate for the total number of marine prokaryotes relies on three recent papers (26–28). The first is a review by Arístegui et al. which included a meta-analysis of values from the literature on the concentration of cells of pelagic prokaryotes at epipelagic (<200 m), mesopelagic (200-1000 m) and bathypelagic (1000-4000 m) depths. Arístegui et al. used many samples of cell density per volume in each depth zone to generate an average cell concentration for each depth zone. Arístegui et al. then generate estimates for the total number of cells per unit area for each depth zone by applying the average cell concentration per unit volume across the entire depth of the zone (≈200 m for epipelagic, ≈800 m for mesopelagic and ≈3000 m for bathypelagic). We used the depth integrated estimates of cell concentration per unit area for each zone in Arístegui et al. (29), and multiplied them by the surface area of the oceans (3.6×10^{14} m^2) to give a total estimate of 1.7×10^{29} cells (link to full calculation). As a second source for estimating the total number of marine prokaryotes, we use Buitenhuis et al. (30). Buitenhuis et al. compiled a database of 39,766 data points consisting of flow cytometric and microscopic measurements of the abundance of marine prokaryotes with observations covering depths even below 4 km. We binned the data along the water column with bins every 100 meters. We calculated the average cell concentrations at each depth bin and multiplied the average cell concentration by the total volume of water in each depth bin to generate an estimate for the total number of marine prokaryotes. We estimate estimate a total of $\approx1.3\times10^{29}$ cells based on the data from Buitenhuis et al. (link to full calculation). The third resource we used for estimating the total number of marine prokaryotes is a recent meta-analysis by Lloyd et al. (28), which gathered ≈20 studies measuring bacterial and archaeal abundance using fluorescent in situ hybridization (FISH). Lloyd et al. fit an equation to predict the concentration of bacteria and archaea based on the depth at which the sample was collected. We used the fits to extrapolate the number of bacteria and archaea across the entire depth of the water column. We then estimated the total number of bacterial and archaeal cells by multiplying the concentration at each depth with the total volume of water at that depth, and integrating across all depths. Lloyd et al. also report the fraction of cells that typically get labeled by the FISH signal. We used the geometric mean of this fraction, which is ≈0.8, to extend the estimate of total bacterial and archaeal cells labelled by FISH to an estimate for the total number of bacterial and archaeal cell in the ocean (link to full calculation). Based on the data from Lloyd et al. we estimate a total of $\approx8\times10^{28}$ cells. As our best guess for the total number of marine prokaryotes, we take the geometric mean of the estimates from Arístegui et al., Buitenhuis et al. and Lloyd et al., which is $\approx1.2\times10^{29}$ cells (link to full calculation). This number corresponds well with the estimate of $\approx1.2\times10^{29}$ cells made by Whitman et al. (25). To estimate the characteristic carbon content of a marine prokaryote, we rely on several studies from the literature (31–35). We use the geometric mean of the values reported, which is ≈11 fg C $cell^{-1}$ as our best estimate for the carbon content of marine prokaryotes (link to full calculation). To generate our best estimate for the total biomass of marine prokaryotes we multiply our best estimate for the total number of marine prokaryotes by our best estimate for the carbon content of marine prokaryotes. We thus

arrive at a total biomass of ≈1.3 Gt C of marine prokaryotes, which is on par with the 2.2 Gt C estimate of Whitman.

We note that these estimates referred to planktonic organisms and did not refer to prokaryotes attached to large particulate organic matter (POM) In general, POM can be divided into two main size fractions - microaggregates (5-500 μm in diameter) and macroaggregates (>500 μm in diameter; 36). Most of the studies on the abundance and relative contribution of particle-attached bacteria and archaea in the oceans are focused on a small number of sampling sites, and thus a robust global accounting of the contribution of particle-attached bacteria and archaea is not easily attainable. We proceed to make a crude estimate of the biomass contribution of bacteria and archaea on microaggregates and macroaggregates. For macroaggregates, we rely on studies which measured the relative fraction of cells attached to macroaggregates out of the total population of cells (37–44). We calculate the average value of those studies, and thus estimate that bacterial and archaeal cells attached to macroaggregates account for ≈3% of the total number of bacterial and archaeal cells in the marine environment (link to full calculation). Our samples of the concentration of bacterial and archaeal cells attached to macroaggregates cover depths up to 1000 meters. We could not find samples of the abundance of particle-attached cells in the bathypelagic realm. We assume that measurements in the epipelagic and mesopelagic realms are characteristic of the bathypelagic realm. To support this assumption, we compare the concentration of macroaggregates measured in the deep-sea (45–47). We find that the concentration of macroaggregates in the deep-sea is similar to the concentration of macroaggregates reported in the studies on which we rely (link to full calculation).

For microaggregates, we have a more limited set of observations (48, 49), but they suggest that bacteria and archaea on microaggregates account for ≈4% of the total number of cells (link to full calculation). It is important to note that from the available data (37, 40, 42, 43, 50) the characteristic volume of bacterial and archaeal cells on aggregates is typically larger than that of free-living bacteria and archaea in the marine environment. The data on the specific factor by which particle-attached cells are larger than free-living cells is very limited. We rely on several studies (37, 40, 42, 43, 50), which suggest that, on average, particle-attached cells contain ≈3-fold more carbon than free-living cells (link to full calculation). This means that even though in terms of cell abundance particle-attached cells account for ≈7% of the total population of cells in the marine environment, in terms of carbon content they account for ≈20% of the total carbon (link to full calculation). In order to account for the additional biomass contributed by particle-attached bacteria and archaea, we use ≈1.6 Gt C (instead of the ≈1.3 Gt C calculated above) as our best estimate for the total biomass of marine bacteria and archaea. We note that depending on the method used to estimate the total number of bacterial cells in the ocean, some cells attached to microaggregates might be counted as free-living (the microaggregates are transparent, so in case cells are counted by microscopy, they might be counted as free-living). Due to the scarcity of data, we assume the distribution of biomass between bacteria and archaea on particles is similar to that of the surrounding water. Our estimates on the total abundance and carbon content of particle-attached cells in the oceans is much less robust that our estimates for free-living cells, and much more comprehensive data is needed to probe the global importance of particle-attached cells in the ocean.

The abundance of archaea in the deep ocean (mesopelagic realm, 200-1000m depth, and bathypelagic realm, 1000-4000m depth) has been reported by Karner et al. (51), which estimated that archaea constitute about half the prokaryotic cells below 1000m in the Hawai'i Ocean Time-series station. Karner et al. (51) used fluorescent *in-situ* hybridization (FISH) with specific 16S probes to estimate the fraction of bacteria and archaea out of the total population of prokaryotes. In shallower waters, bacteria dominate the population (51). Similar findings were found using FISH in the Atlantic Ocean (52, 53). To rigorously estimate the fraction of archaea out of the total biomass of marine prokaryotes, we rely on two independent methods - FISH and 16S rDNA sequencing. A recent paper by Lloyd et al. (28) collected studies reporting on the number of archaea and bacteria at different depths, based on FISH. As part of the procedure to estimate the total number of marine prokaryotes from the data in Lloyd, we calculate the total number of archaeal and bacterial cells. The fraction of archaeal cells out of the total number of bacterial and archaeal cells is ≈20%. In the epipelagic, mesopelagic and bathypelagic zone, archaea represent ≈6%, 24% and 35% of the the total number of cells, respectively (link to full calculation). An alternative methodology to quantify the fraction of archaea out of the marine prokaryote population is by using 16S rDNA sequencing. We use studies which have measured the fraction of archaea in the epipelagic, mesopelagic and bathypelagic realms. For the epipelagic and mesopelagic realms, we use data from the Tara Oceans campaign (54), which is based on ≈250 samples worldwide. The fraction of archaeal 16S rDNA sequences out of the total pool of 16S sequences in the epipelagic and mesopelagic realms is ≈4% and ≈14%, respectively (link to full calculation). For the bathypelagic realm, we rely on data from the recent Malaspina campaign (55), which is based on 30 samples ranging in depth from ≈2 km to ≈4 km. The average fraction of archaeal 16S rDNA sequences out of the total pool of 16S sequences is ≈15% (link to full calculation). Estimates based on 16S rDNA sequencing data are lower by about 2-fold than FISH-based estimates for the fraction of archaea across depths. This might be caused by the fact that the copy number of rRNA operons in bacterial genomes is on average ≈2-fold larger than that of archaeal genomes (56). We use the geometric mean of estimates from the two methodologies as our best guess for the fraction of archaea out of the total population of marine prokaryotes in the different layers of the ocean. Our best estimates for the epipelagic, mesopelagic and bathypelagic realms are thus ≈7%, 26%, and 32%, respectively (link to full calculation). From the data in Arístegui et al. and in Lloyd et al. we generate estimates for the fraction of prokaryotes that are found in the epipelagic, mesopelagic, and bathypelagic realms. Applying the fractions of archaea across the different environments, we estimate that archaeal cells represent ≈20% of the total cells of marine prokaryotes (link to full calculation). As the average cell sizes of bacteria and archaea don't seem to vary considerably (57), we estimate that the biomass of marine archaea represents ≈20% of the total biomass of marine prokaryotes, which is **≈0.3 Gt C**. This puts the estimate of the biomass of marine bacteria at **≈1.3 Gt C** (link to full calculation).

We now analyze the associated uncertainty of the estimate for the total biomass of marine bacteria and archaea, which we report as a fold-change factor from the mean representing a range akin to a 95% confidence interval of the estimate. In this analysis, we consider the following factors. First, we assess the uncertainty associated with the estimate of the total number of marine prokaryotes. We then estimate the uncertainty associated with the estimate of the characteristic

carbon content of a single marine prokaryote. Finally, we assess the uncertainty associated with the estimate of the fraction of archaea out of the total population of marine prokaryotes. The intra-study uncertainty reported in Arístegui et al. for the estimate of the cell concentration is ≈10% of the mean. Buitenhuis et al. and Lloyd et al. do not report uncertainty ranges for the estimate of the total number of cells of marine prokaryotes. The inter-study uncertainty between these three studies is ≈1.5-fold (link to full calculation). For estimating the characteristic carbon content of a single marine prokaryote, we used 9 independent studies which measure carbon content in the open ocean. The maximum reported intra-study uncertainty is ≈1.4-fold, and the inter-study uncertainty between the five studies is ≈1.4-fold (link to full calculation). We thus chose to use ≈1.4-fold to project the uncertainty associated with the carbon content of a single marine prokaryote. Combining the uncertainty associated with our estimate of total number and the uncertainty associated with our estimate of the carbon content of a single marine prokaryote, we arrive at an uncertainty of ≈1.7-fold for the total biomass of marine prokaryotes. We also analyze the uncertainty associated with our estimate of the fraction of the total biomass of marine bacteria and archaea which is particle-attached. Our estimate relies on two main factors - the fraction of the total number of cells which is particle-attached, and the carbon content of particle-attached bacteria and archaea relative to free-living bacteria and archaea. Our projection for the uncertainty associated with our estimate of the fraction of the total number of cells which is particle-attached is based on collecting the intra-study and inter-study uncertainty associated with the estimate of the fraction of the total number of cells which are particle-attached, and using the maximum of the uncertainties values, which is ≈3-fold (link to full calculation). We repeat the same procedure for the estimate of the carbon content of particle-attached cells relative to free-living cells, and project an uncertainty of ≈3-fold (link to full calculation). We combine the uncertainties associated with the two factors and project an uncertainty of ≈5-fold associated with our estimate of the total biomass of marine bacteria and archaea which are attached to particles (link to full calculation). We then combine the uncertainty of the total biomass of free-living bacteria and archaea and particle-attached bacteria and archaea, and project an uncertainty of ≈1.8-fold associated with our estimate of the total biomass of marine bacteria and archaea.
For the fraction of archaea out of the population of marine prokaryotes, the intra-study uncertainty is ≈1.2-fold for Salazar et al. and ≈1.1-fold for Lloyd et al. (link to full calculation). The inter-study uncertainty between FISH-based studies and 16S rDNA sequencing-based studies is around ≈2.3-fold for archaea and ≈1.3-fold for bacteria (link to full calculation). We use the higher inter-study uncertainties for projecting the uncertainty of the fraction of marine archaea and bacteria out of the total marine prokaryote population. Combining the uncertainties for the biomass of marine prokaryotes with the uncertainties associated with the estimate of the fraction of archaeal and bacterial cells out of the total population of marine prokaryotes, we project an uncertainty of ≈2-fold for the biomass of marine bacteria, and ≈3-fold for marine archaea (link to full calculation).

Soil

To estimate the total biomass of soil bacteria and archaea, we rely on the estimate of the total biomass of soil microbes we derived in the soil fungi section. We estimate a total microbial biomass of 20 Gt C, of which ≈12 Gt C are fungal. This leaves us with ≈8 Gt C of bacterial and

archaeal biomass. In order to derive the respective fractions of archaea and bacteria out of this total biomass, we rely on four independent methods which estimate the fraction of archaea out of the total biomass of bacteria and archaea. The methods we rely upon are 16S rDNA sequencing, 16S rDNA quantitative PCR (qPCR), fluorescent *in-situ* hybridization (FISH), and catalyzed reporter deposition FISH (CARD-FISH). We calculated the mean estimate of the fraction of archaea out of the total biomass of soil bacteria and archaea for each method. We then used the geometric mean of values from the different methods as our best estimate for the fraction of archaea out of the total biomass of soil bacteria and archaea. For our 16S rDNA sequencing-based estimate, we rely on a study which reported values for the fraction of archaea out of the total population of soil bacteria and archaea in 146 soils from across the globe (58). The mean fraction of archaea reported by Bates et al. (58) is ≈2%. We account for the lower rRNA operon copy number in archaea (56) by multiplying the measured fractions by a factor of 2. This procedure does not affect our results significantly. For our 16S qPCR-based estimate, we rely on a recent study which reported the fraction of archaea out of the total population of soil bacteria and archaea in grasslands, forests and croplands in Korea (59). The mean fraction of archaea reported by Hong & Cho (59) is ≈3%. For our FISH-based estimate, we assembled data from about 10 studies (60–68). We calculate the mean fraction of archaea across these studies, which is ≈20% as our best FISH-based estimate for the fraction of archaea out of the total population of soil bacteria and archaea (link to full calculation). For our CARD-FISH-based estimate, we assembled data from four studies (69–72). We calculate the mean fraction of archaea across these studies, which is ≈20% as our best CARD-FISH-based estimate for the fraction of archaea out of the total population of soil bacteria and archaea (link to full calculation). To generate our best estimate for the fraction of archaea out of the total biomass of soil bacteria and archaea, we use the geometric mean of these four estimates based on the four different methodologies. We thus estimate that archaea represent ≈7% of the total biomass of soil bacteria and archaea (link to full calculation). We combine this estimate with our estimate for the total biomass of soil bacteria and archaea, which is 8 Gt C, and arrive at an estimate of **≈0.5 Gt C** of soil archaea, and **≈7 Gt C** of soil bacteria (link to full calculation).

The different methods we rely on to estimate the fraction of archaea out of the population of soil bacteria and archaea have various caveats associated with them. In general, for methods based on quantifying relative abundances of 16S rDNA sequences, we rely on the assumption that the abundance of 16S rDNA sequences is proportional to the biomass of bacteria and archaea. Relying on 16S sequence abundance as a proxy for biomass is not a well-established practice. For qPCR, a recent meta-analysis claimed relative fractions of archaea and bacteria calculated using this methodology are reliable for the marine environment and in subseafloor sediments (28). However, we only have a limited amount of data which is based on qPCR measurements. The same meta-analysis by Lloyd et al. also states that the use of FISH or CARD-FISH for quantifying the abundance of archaea and bacteria is very sensitive to the details of the experimental protocol and was found to reliably represent the community structure of subseafloor sediments only under a specific set of protocol parameters (e.g. CARD-FISH with cell permeabilization using proteinase K). It is thus not obvious that the estimates based on FISH or CARD-FISH represent reliably the fraction of archaea out of the total population of soil bacteria and archaea. Due to lack of better options, we chose to use the geometric mean of these four

methods, each with its own caveats as our best estimate. We hope that further research will shed light on the appropriate methodology to quantitatively describe biomass distribution of soil microbes.

We now analyze the associated uncertainty of the estimate for the total biomass of soil bacteria and archaea, which we report as a fold-change factor from the mean, representing a range akin to a 95% confidence interval of the estimate. We first assess the uncertainty associated with our estimate of the fraction of archaea out of the total biomass of soil bacteria and archaea. As a measure of the uncertainty associated with our estimate of the fraction of archaea out of the total biomass of soil bacteria and archaea, we collect the intra-study, inter-study, inter-habitat and inter-method uncertainties within and between each of the four methods we rely on to estimate the fraction of archaea out of the total biomass of soil bacteria and archaea. We use the maximal uncertainty among this collection as our best projection of the uncertainty associated with our estimate of the fraction of archaea out of the total biomass of soil bacteria and archaea. We thus project that ≈4-fold uncertainty associated with our estimate of the fraction of archaea and ≈1.5-fold uncertainty associated with our estimate of the fraction of bacteria out of the total biomass of soil bacteria and archaea (link to full calculation). We combine this uncertainty with the uncertainty associated with our estimate of the total biomass of soil prokaryotes (bacteria and archaea), which we derive in the soil fungi section. The uncertainty we project as associated with the total biomass of soil prokaryotes is ≈4-fold. We thus project an uncertainty of ≈6-fold for our estimate of the total biomass of soil archaea, and ≈4-fold for our estimate of soil bacteria (link to full calculation).

Marine deep subsurface sediment

The two major habitats of microbes in the marine deep subsurface are subseafloor sediments and the oceanic crust (73, 74). We first focus on subseafloor sediments, as much more data is available on this environment. We then turn to look at the oceanic crust.
Whitman et al. (25) originally estimated the global biomass of bacteria and archaea in subseafloor sediments to be around 300 Gt C residing within $\approx 3\times10^{30}$ cells. A later study (75) revealed that the original estimates by Whitman were based on extrapolation from samples which were not representative of the different levels of productivity in the marine environment. Using additional sampling, the study by Kallmeyer et al. updated the estimate for the total number of prokaryotes in the subseafloor sediments to around $\approx 3\times10^{29}$ cells, an order of magnitude lower than Whitman's original estimate. The estimate by Kallmeyer et al. is based on sampling of cell densities worldwide at different depths. Kallmeyer et al. built a model to predict cell concentration as a function of location and depth below seafloor. The model uses distance from shore and sedimentation rate as the primary explanatory variables. The specific parameters of the model are given in detail in (75). Kallmeyer et al. plugged into the model global data on sedimentation rates and the distance from shore, and used the model to predict cell concentrations at each location in the marine subsurface. Kallmeyer et al. then integrated the predicted cell concentrations across the entire volume subseafloor sediments the generate an estimate for the total number of cells in subseafloor sediments. A later study (76), gives a slightly higher estimate for the total number of prokaryotic cells in subseafloor sediments. Parkes et al. calculated the total number of cells in the subseafloor sediments by using regression of cell concentrations

across different sampling points by depth, generating an average cell concentration for each depth and multiplying the average cell concentration by the total global volume at each depth. In total, Parkes et al. estimate $\approx 5\times10^{29}$ cells. Parkes et al. claim that the difference between their estimate of $\approx 5\times10^{29}$ cells and the $\approx 3\times10^{29}$ cells estimated by Kallmeyer et al. could stem from areas with higher activity of deep sediment prokaryotes that were included in their study and excluded by Kallmeyer et al. As our best estimate for the number of cells in the marine deep subsurface, we use the geometric mean of the estimates from Kallmeyer et al. and Parkes et al., which is $\approx 4\times10^{29}$ cells (link to full calculation). To estimate the total biomass of prokaryotes in subseafloor sediments we need to estimate a characteristic carbon content for an average single prokaryote in subseafloor sediments. For estimating the carbon content of a single prokaryote in subseafloor sediments, we rely on four previous studies. The first study to estimate the carbon content of prokaryotes in subseafloor sediments is Parkes et al. (77), which measured a characteristic cell volume of $\approx 0.2\ \mu m^3$. Parkes et al. converted this volume into carbon content by using a carbon density per unit volume of 310 fg C μm^{-3} (78). Parkes et al. thus arrived at an estimate of ≈65 fg C $cell^{-1}$. A later study by Whitman et al. (25) relied on a single study which measured cell dry weight from a terrestrial aquifer to arrive at a carbon content of ≈86 fg C. In one of the main studies establishing the important role of archaea in subseafloor sediments, Lipp et al. (79) used measures for the characteristic diameter of a cells of ≈0.5 μm, based on data from Peru collected by Biddle et al. (80). They assumed cells as spherical, and thus arrived at an estimate of ≈0.0625 μm^3 for the volume of cells. Lipp et al. converted the characteristic volume of cells to carbon content based an allometric equation constructed for marine bacteria (81), which have similar volumes. Lipp et al. arrived at an estimate of ≈18 fg C $cell^{-1}$. In their study, Kallmeyer et al. (75) used microscopy to measure the length and width of microbes in subseafloor sediments at the South Pacific Gyre. Kallmeyer et al. then estimate the biovolume of cells assuming either rod or spherical shapes and converted biovolume to carbon content using the same allometric equation from Simon & Azam (81). Kallmeyer et al. calculated a lower carbon content in each cell of ≈14 fg $cell^{-1}$. A recent study (82) measured cell volumes as well as amino acid content per cell in samples from the Baltic sea. Braun et al. used the assumption that carbon in amino acids constitute ≈55% of cellular carbon (83) to estimate the carbon content of cells. The estimate from Braun et al. of ≈22 fg C $cell^{-1}$.

These four studies represent two independent methods for estimating carbon content (based on cell biovolume and based on amino acid carbon measurement), and four independent sampling efforts. To generate our best estimate for the carbon content of prokaryotes in subseafloor sediments we calculate the mean carbon content for each of the two methodologies. To calculate the mean carbon content for the biovolume-based methodology, we use reported biovolumes from the four studies. To convert biovolume to carbon content, we use two independent methods. The first method is by using the allometric relation from Simon & Azam, and the second is by multiplying by carbon density of 310 fg C μm^{-3} based on Fry et al. (78). For each of those conversion methods, we calculate the geometric mean across studies. This gives us two characteristic carbon contents, which are based on biovolume measurements. Our estimate for the amino acid-based carbon content is based on the mean value measured in Braun et al. across samples. These three values for the carbon content of prokaryotes in subseafloor sediments are independent of each other and are based on different types of assumptions. We use the geometric mean of the biovolume-based carbon content estimate and the amino acid-based carbon content

as our best estimate for the carbon content of prokaryotes in subseafloor sediments. Our best estimate following this methodology is ≈24 fg C $cell^{-1}$ (link to full calculation). Multiplying our best estimate for the number of cells in subseafloor sediments of $\approx 4 \times 10^{29}$ cells by our best estimate for the carbon content of prokaryotes in subseafloor sediments of ≈24 fg C, we arrive at our best guess for the total biomass of prokaryotes in subseafloor sediments of ≈10 Gt C. The combination of lower estimates for the total number of cells in the subseafloor sediments, as well as the lower estimate of carbon content per cell results in an estimate which is ≈30-fold lower than the original estimate by Whitman. The need for such a significant revision was already pointed out by Kallmeyer et al. (75).

In 2008, a study (79) estimated that the biomass of prokaryotes in subseafloor sediments is dominated by archaea, based on the fraction of intact polar membrane lipids (IPLs) that is associated with archaea. However, later studies (84) have discovered that due to a low degradation rate of IPLs, using IPLs as a measure of archaeal biomass will overestimate their biomass. The current view which integrates independent methods such as IPLs, FISH and 16S sequencing, is that archaea and bacteria constitute a similar fraction of the total number of prokaryotes (28, 76, 85). Our best estimate for the fraction of archaea out of the population of prokaryotes in subseafloor sediments is based on two independent methods - qPCR and catalyzed reporter deposition FISH (CARD-FISH). These two methods were recently validated in a meta-analysis to be reliable for reporting the fraction of archaeal cells out of the total population of prokaryotes in subseafloor sediments (28). We use the values reported in Lloyd et al. for sediments which are deeper than 10 cm (the range for which biomass is reported in Kallmeyer et al.). For each methodology we calculate a mean fraction across all sampled depths. For our best estimate, we use the geometric mean of the two fractions - the one based on qPCR and CARD-FISH. The geometric mean of the two methods is ≈1/3 (link to full calculation). Cell sizes of archaea and bacteria in subseafloor sediments are similar (75, 86), and thus we estimate that the biomass of archaea in subseafloor sediments is roughly one third of the total biomass of prokaryotes, or **≈3 Gt C**. Correspondingly, our best estimate for the biomass of bacteria in subseafloor sediments is **≈7 Gt C** (link to full calculation).

The estimate by Kallmeyer et al. does not take into account biomass of bacteria and archaea in the top 10 cm of subseafloor sediments. To estimate the biomass contribution of bacteria and archaea in the top 10 cm of sediments, we rely on two main sources which report the total biomass of bacteria and archaea in the top layer of sediments (87, 88). There is a large discrepancy between the values reported in these two studies. Whereas Wei et al. (87) reports ≈0.04 Gt C of bacteria and archaea in the top ≈10 cm of subseafloor sediments, Danovaro et al. (88) reports ≈1.5 Gt C in the top 50 cm of subseafloor sediments. This discrepancy is present even though both studies rely in large part on the same data source (89). If we use the exponential decay in cell abundance with sediment depth reported in Danovaro et al. to estimate the biomass of bacteria and archaea only in the top 10 cm of the sediment from Danovaro et al., we arrive at ≈0.6 Gt C. This order of magnitude difference between the two estimates stems in small part from different values used for the characteristic carbon content of bacteria and archaea. Even if we take the higher estimate of ≈0.6 Gt C, it would not affect our total estimate for the biomass of bacteria and archaea in the subseafloor sediment significantly.

In addition to subseafloor sediments, microbes are also present in fluids inside the oceanic crust. This habitat is much less explored than subseafloor sediments, and thus estimates of the total biomass of prokaryotes in the ocean's crust are more speculative. The total volume of the oceanic crust with temperatures that can sustain life (<120° C) is ~10^{18} m^3 (90). This includes three main layers - a basalt layer that is 0.5-1-kilometer-thick, a layer of sheeted dike complex and a bottom layer of gabbroic rock. The basaltic layer is the only layer to host significant porosity and permeability and hence this is probably the most significant portion to consider for microbiological habitation and processes (74). The total volume of pore water in the basaltic layer of the oceanic crust is estimated at ≈2×10^{22} mL (91). A very sparse collection of measurements of cell concentrations is available for the oceanic crust. The available data suggest cell concentrations on the order of ~10^4 cells mL^{-1} in crustal fluids. These concentrations of cells were observed both in warm fluids of the eastern flank of the Juan de Fuca Ridge (92, 93), and the cooler fluids of the North Pond site which are supposed to be more representative of the upper crust (94). A major gap in understanding the abundance of microbes in the oceanic crust is that most of the measurements of microbial biomass are for crustal fluids, and there are very few samples of microbial populations attached to the rocks themselves. Assuming oceanic crust aquifers are similar in nature to deep terrestrial aquifers, cells present in oceanic crustal fluids represent only a small minority of the cells in the oceanic crust, with ≈10^2-10^3 more attached cells than free-living cells. This claim might be an overestimate, as complete basalt samples, and not crustal fluids samples, taken from the same North Pond site, contained as a whole ~10^4 cells per cm^3 of rock (95). The extraction procedure of Zhang et al. is supposed to detach cells attached to surfaces, so these results suggest that the total ratio between attached and unattached cells might be much lower than in the terrestrial aquifer. Still, even this procedure might not extract the entire population of attached cells. For estimating the total biomass of prokaryotes in the oceanic crust, we chose to use ratios of attached/unattached cells of ~10^2-10^3 (serving potentially as a liberal upper limit), which brings the total cell densities to ~10^6-10^7 cells mL^{-1}. Applying these total cell densities across the entire volume of the ocean crust aquifer, we get an estimate of a total of 2×10^{28}-2×10^{29} cells. Assuming prokaryotes in the oceanic crust have similar carbon content to those present in subseafloor sediments or in the terrestrial deep subsurface, we assume a carbon content of ≈25 fg C per cell. Multiplying the total number of cells by the characteristic carbon content of a bacterial or archaeal cell, we estimate a total biomass of prokaryotes in the oceanic crust of 0.5-5 Gt C. As our best estimate for the biomass of prokaryotes in the oceanic crust, we use the geometric mean of the higher and lower estimates, which is ≈1.5 Gt C. Cells might also be present in lower layers of the ocean crust, such as the gabbroic layer. However, from the limited available data, cell densities seem to be more than an order of magnitude lower than those found in the upper crust (96), and thus will not affect the estimates for the total biomass of microbes in the ocean crust significantly. By this estimate, the global biomass of prokaryotes in the oceanic crust represents ≈15% of the total biomass of prokaryotes in subseafloor sediments. Therefore, including it or excluding it will not dramatically affect our estimate for the total biomass of prokaryotes in the marine deep subsurface. Notably, our estimate of ≈1.5 Gt C is markedly lower than an earlier estimate by Heberling et al. (90), which estimated about 200 Gt C. The specific procedure Heberling et al. used to estimate the total biomass of prokaryotes in the oceanic crust is not available, but it involved estimating the total volume of pore water in the

oceanic crust and estimating the total volume of bacterial and archaeal cells using the assumption that cells account for ≈0.016% of the pore water volume. This value for the fraction of cells out of the total volume of pore water is based on Whitman et al. (25), which used data in Harvey et al. (97). Harvey et al. measured a total amount of ≈2-4×10^7 cells per cm^3 of aquifer material. Whitman et al. presumably arrived at the estimate of ≈0.016% by assuming a volume of a bacterial cell is ≈1 μm^3 and using porosity data from Harvey et al. of ≈0.15-0.19. The volume of a bacterial cell used by Whitman et al. ≈1μm^3 is very likely a large overestimate, as this volume is usually characteristic of *Escherichia coli* cells grown in the lab (Bionumber ID 109483). In deep and nutrient-limited environments, such as subseafloor sediments and the terrestrial deep subsurface, cells are much smaller, with a characteristic volume of ≈0.1 μm^3. This means the fraction of pore water taken up by cells is an order of magnitude lower than the fraction estimated by Whitman et al. In addition, Whitman et al. calculated the fraction of pore water colonized by cells based on samples which are pretty shallow (<50 m in depth), and contain free-living cell densities of ~10^6 cells mL^{-1}, about two-orders of magnitude larger than the ones measured in the oceanic crust. As both in the terrestrial deep subsurface and in subseafloor sediments an exponential decrease in cell densities with sample depth (75, 98) was observed, using shallow samples to estimate the number of cells across the entire volume of the oceanic crust is probably an overestimate. Overall, we believe these factors help to explain the discrepancy between our estimate and the estimate provided in Heberling et al.

We now analyze the associated uncertainty of the estimate for the total biomass of bacteria and archaea in the marine deep subsurface, which we report as a fold-change factor from the mean representing a range akin to a 95% confidence interval of the estimate. The two main factors affecting the estimate of the biomass of marine deep subsurface prokaryotes are the number of cells and the carbon content of each cell. For the number of cells, the intra-study uncertainty reported in Parkes et al. (76) is ≈5-fold for the 95% confidence interval. Kallmeyer et al. (75) reports a standard deviation of ≈2.6-fold for the total number of prokaryotes in the deep marine subsurface, based on bootstrapping the model parameters. We assume this bootstrapping results in a distribution which can be approximated as a lognormal distribution, and thus the 95% confidence interval for the estimate is ≈7-fold range (link to full calculation). The inter-study uncertainty between the number of cells reported in Parkes et al. and Kallmeyer et al. is ≈2-fold (link to full calculation). We thus choose to take the highest uncertainty range out of the intra-study and interstudy uncertainties and project an uncertainty of ≈7-fold for the number of prokaryotes in the marine deep subsurface. For the uncertainty of the carbon content of marine deep subsurface prokaryotes, the studies on which we rely do not report uncertainty ranges around the mean carbon content but do supply the ranges of measured values. Thus, to project the uncertainty associated with the estimate of the carbon content of marine deep subsurface prokaryotes, we calculate the inter-study uncertainty between studies using the same methodology, and between different methodologies. The highest uncertainty amongst this collection of uncertainties is ≈2-fold (link to full calculation). We thus use this uncertainty as our projected uncertainty associated with the estimate of the carbon content of marine deep subsurface prokaryotes. Combining the uncertainty of the carbon content of cells in the marine deep subsurface with the uncertainty of the total number of cells, we project that the uncertainty of the biomass of marine deep subsurface prokaryotes is about 8-fold. In addition to the

uncertainty associated with the estimate of the total biomass of the marine deep subsurface prokaryotes, estimating the fraction of biomass of archaea has uncertainty associated with it. To estimate the uncertainty associated with estimates for the fraction of archaea out of the total population of prokaryotes, we rely on the data from Lloyd et al. (28), who used two independent methods to estimate the fraction of archaea out of the population of prokaryotes. We estimate the uncertainty of the data within each method, as well as the uncertainty between the characteristic values for each method. Whereas the intra-method uncertainty is about 1.2-fold, the uncertainty between methods is ≈1.6-fold for the fraction of archaea and ≈1.3-fold for the fraction of bacteria (link to full calculation). As the cell size of archaea and bacteria in the marine deep subsurface is similar (75, 86), we estimate that it does not introduce further uncertainty to the total estimate. We combine the uncertainty of the fraction of archaea and bacteria out of the total biomass of prokaryotes with the uncertainty of about an order of magnitude for the total biomass of prokaryotes in the marine deep subsurface. We thus estimate that the total uncertainty of the estimate for the biomass of marine deep subsurface archaea and bacteria is roughly 8-fold each (link to full calculation).

Terrestrial deep subsurface

We follow the definition made by Whitman et al. for the terrestrial deep subsurface as the terrestrial substratum deeper than 8 m, excluding soil. In the spirit of Kallmeyer et al. (75) revisiting the question of the marine sedimentary prokaryote biomass, McMahon & Parnell (98) revisited Whitman's original estimate of the terrestrial sedimentary prokaryote biomass (25). The original estimate by Whitman was based on three separate approaches. For the lower limit, Whitman et al. used sampling of cell concentrations from unconsolidated sediments and inferred the total number of cells globally. As unconsolidated sediments represent only ≈20% of the terrestrial deep subsurface (25), this estimate was considered as an underestimate. For the upper limit, Whitman et al. used sampling of cell concentrations from groundwater, and extrapolated the global biomass by using the ratio of attached and unattached cells in groundwater and the total groundwater volume. Alternatively, Whitman et al. used the soil porosity and the fraction of pore volume occupied by cells. From the estimate of the total amount of cells in the terrestrial deep subsurface, Whitman et al. estimated the global biomass of prokaryotes in the terrestrial deep subsurface by using a carbon content of 86 fg cell^{-1}. More recently, McMahon & Parnell (98), used a similar approach of extrapolating groundwater cell concentrations along with the ratio of attached to unattached cells. Our estimate of the total biomass of prokaryotes in the terrestrial deep subsurface relies on data from McMahon & Parnell, along with an updated estimate for the total volume of groundwater from Gleeson et al. (99). We use the measurements of cell concentrations in groundwater reported in McMahon & Parnell and followed a similar procedure of binning the measurements along depth bins of 250 meters. In each depth bin, we calculate the characteristic concentration of cells per mL of groundwater. We generated two types characteristic concentrations at each depth bin - an estimate which uses the arithmetic mean of cell concentrations at each depth bin, and an estimate which uses the geometric mean of cell concentrations at each depth bin. The estimate based on the arithmetic mean is more susceptible to sampling bias, as even a single measurement which is not characteristic of the global population (such as samples which are contaminated with organic carbon sources, or samples

which have some technical biases associated with them) might shift the average concentration significantly. On the other hand, the estimate based on the geometric mean might underestimate global biomass as it will reduce the effect of biologically relevant high biomass concentrations. As a compromise between these two caveats, we chose to use as our best estimate the geometric mean of the estimates from the two methodologies. Cell concentrations in depth bins with missing values were extrapolated from a fit to an exponential decay function of the available data. To estimate the total number of cells of prokaryotes in groundwater, we estimate the total volume of water at each depth bin from data reported in Gleeson et al. We multiply the characteristic cell concentration at each depth by the total volume of groundwater at each depth, and sum over all depth bins to estimate the total number of prokaryotes in groundwater. From the total number of cells in groundwater, we estimate the total number of cells in the terrestrial deep subsurface by following the same procedure used by McMahon & Parnell. McMahon & Parnell use the ratio of attached to unattached cells and estimate a range of 10^2-10^3 for this ratio. As our best estimate for the ratio of attached to unattached cells in the terrestrial deep subsurface, we use the geometric mean of these minimum and maximum values from the range reported by McMahon & Parnell. We multiply our estimate for the total number of cells in groundwater by our best estimate for the ratio of attached to unattached cells, and arrive at an estimate of $\approx 2\times10^{30}$ prokaryotes in the terrestrial deep subsurface (link to full calculation). To get from the total number of prokaryotes to an estimate of the total biomass of prokaryotes in the terrestrial deep subsurface, we use the carbon content of single bacterial or archaeal cells used by McMahon & Parnell, which is ≈ 26 fg C $cell^{-1}$. This value is similar to our best estimate of the characteristic carbon content of cells in subseafloor sediments. In total, we estimate a biomass of ≈60 Gt C for prokaryotes (bacteria and archaea) in the terrestrial deep subsurface (link to full calculation). Notably, this estimate only takes into account prokaryotes present in the top 2 km of the terrestrial subsurface. The top ≈5 km of the terrestrial subsurface are estimated to be habitable by life (100). However, using the regression of cell concentration with depth, the average cell concentrations in depths of 2-5 km below the surface are estimated to be more than two orders of magnitude lower than the average cell concentrations in the upper 2 km. Thus, the biomass of bacteria and archaea in depths below 2 km probably do not affect the total estimate significantly. Nevertheless, the trends observed in the top 2 km of the terrestrial subsurface might not be applicable to deeper layers, and further research is needed to establish the actual contribution from deeper layers.

There are only few studies about the abundance of archaea in the terrestrial subsurface, and thus estimating accurately the fraction of the biomass of terrestrial deep subsurface prokaryotes contributed by archaea is challenging. To estimate the fraction of archaea out of the total biomass of prokaryotes in the terrestrial subsurface, we rely on four different lines of evidence. The first three lines of evidence are based on three different methodologies, namely 16S amplicon sequencing (101–104), qPCR of rDNA (105–107), and Fluorescent *in situ* Hybridization (FISH; 108). For qPCR, a recent meta-analysis claimed relative fractions of archaea and bacteria calculated using the methodology are reliable (28). For 16S sequencing data, we account for the lower rRNA operon copy number in archaea (56) by multiplying the measured fractions by a factor of 2. This procedure does not affect our results significantly. Due to the scarcity of data regarding the biomass of archaea in the terrestrial deep subsurface, we chose to also use as an additional line of evidence values for the abundance of archaea in another subsurface

environment, the marine deep subsurface, where archaea constitute ≈35% of the biomass of prokaryotes (28). To generate our best estimate for the abundance of archaea, we integrate the available data to get a characteristic value for each line of evidence. We note that the values vary widely between 0.004% and 75%. We use the geometric mean of the characteristic values from all the different lines of evidence. Our best estimate for the fraction of biomass of archaea out of the population of prokaryotes in the terrestrial subsurface is ≈6% (link to full calculation). Combining this estimate with our best estimate for the biomass of prokaryotes in the terrestrial subsurface, we get an estimate for the total biomass of archaea in the terrestrial deep subsurface of **≈4 Gt C**. Correspondingly, our best estimate for the biomass of terrestrial deep subsurface bacteria is **≈60 Gt C** (link to full calculation).

We now analyze the associated uncertainty of the estimate for the total biomass of bacteria and archaea in the terrestrial subsurface, which we report as a fold change factor from the mean representing a range akin to a 95% confidence interval of the estimate. Assessing the uncertainty associated with the estimate of the biomass of the terrestrial deep subsurface prokaryotes is challenging due to the scarcity of data sources and the lack of independent estimates for the total biomass of terrestrial deep subsurface prokaryotes. We try and quantify the uncertainty associated with each of the factors leading to the final estimate. One factor which introduces uncertainty into the estimate is the procedure of binning measurements along depth and calculating a characteristic concentration of cells at each depth bin. A second parameter controlling the estimate of the terrestrial deep subsurface prokaryote biomass is the volume of aquifers present in each depth. The volume of aquifers present in each depth range is dependent on the total volume of global aquifers, which is estimated at $\approx 2\times10^{22}$ mL (99), along with the characteristic porosity of the soil, which decreases with depth and thus affects the amount of aquifer water present at each depth. Cells in aquifers are divided to unattached cells and attached to soil grains. A third main parameter is the ratio of attached to unattached cells. A fourth parameter controlling the estimate is the average carbon content per cell. As a measure of the uncertainty associated with the procedure of calculating characteristic concentration of cells at each depth bin, we look at the uncertainty of the average cell concentration at each depth bin. We also calculate the uncertainty of the geometric mean of the estimates generated by our two methodologies for calculating the average cell concentration at each depth bin. We take the maximum uncertainty of these different values, which is ≈2.3-fold (link to full calculation). For the total volume of groundwater, Gleeson et al. (99) reports a range of $1.6\text{-}3\times10^{22}$ mL of groundwater down to 2 km. This range represents the standard error range, so we estimate the 95% confidence interval for the volume of groundwater is ≈2-fold (link to full calculation). For the attached to unattached ratio, McMahon & Parnell report a range of $10^2\text{-}10^3$ (the vast majority is attached). As McMahon & Parnell report a range for the ratio of attached to unattached cells, it is hard to quantify rigorously the uncertainty associated with this range. In our estimate we use the geometric mean of this ratio range, so we estimate the associated uncertainty of this parameter to be around 10-fold (link to full calculation). For the uncertainty of the carbon content of cells in the terrestrial deep subsurface, we rely on the uncertainty we projected for the carbon content of cells in subseafloor sediments (see marine deep subsurface prokaryote section), which is ≈2-fold (link to full calculation). Combining all these different sources of uncertainty, we calculate ≈14-fold uncertainty associated with the biomass of prokaryotes in the terrestrial deep subsurface (link to

full calculation). There are other sources of uncertainty for which we are not able to provide a quantitative estimate. The procedure of binning cell concentrations with depth and fitting an equation which extrapolates cell concentrations across all depths has uncertainty associated with it, but this uncertainty is hard to quantify. The uncertainty stemming from excluding the contribution from groundwater deeper than 2 km is also hard to quantify, as the cell concentration at those depths and the volume of groundwater are poorly characterized. Considering these additional uncertainties, we project the uncertainty of biomass of terrestrial prokaryotes to be about 20-fold. In addition to the uncertainty associated with the estimate of the total biomass of the terrestrial deep subsurface prokaryotes, estimating the fraction of biomass of archaea has uncertainty associated with it. We rely on the variability between different methodologies in the characteristic values for the fraction of archaea out of the population of prokaryotes in the terrestrial deep subsurface, and project an uncertainty of ≈17-fold for the fraction of archaea, and ≈1.5-fold for the fraction of bacteria (link to full calculation). We assume that the distribution of cell sizes for terrestrial deep subsurface archaea is similar to that of the marine deep subsurface, and that the cell size of archaea and bacteria are similar (75, 86). Combining the uncertainty in the fraction of archaea with the uncertainty of the biomass of prokaryotes in the terrestrial deep subsurface, we get an uncertainty estimate of about 60-fold and 20-fold for the biomass of archaea and bacteria in the terrestrial deep subsurface, respectively (link to full calculation).

Fungi

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

The estimate of the global fungal biomass is based on two contributions, one from soil fungi, and the second from plant associated mycorrhizal symbionts. We also discuss possible contributions from other environments such as the ocean or subsurface in a dedicated section below.

Soil fungi

In order to estimate the global soil fungal biomass, we start with the more general category of microbial biomass, which includes both fungi and other microbes such as prokaryotes. We then estimate the fraction of the total microbial biomass that is contributed by fungi. Estimates in Whitman et al. (25) for the total biomass of soil bacteria and archaea (not including soil fungi) are based on prokaryotic densities in forest soils and in other soils. For forests, a density of 4×10^7 cells per gram soil in the top 1 meter was used, and a density of 10^6 cells per gram soil was used for 1-8 meters depth. For other soils a value of 2×10^9 cells per gram for the top 1 meter and 10^8 cell per gram for 1-8 meters depth were used. Based on these numbers, Whitman arrived at an estimate of 26 Gt C for the soil prokaryote biomass. The estimate by Whitman et al. relies on very local data to extrapolate the total biomass of soil microbes. Other earlier estimates such as that of Wardle et al. (109) also rely on limited data compared to recent studies. Therefore, we chose to use more recent and comprehensive studies as the basis for our estimate of the total biomass of soil microbes.

Recently two comprehensive studies by Xu et al. (110) and Serna-Chavez et al. (111) have estimated the total biomass of soil microbes. The first study, by Xu et al. (110), compiled a data set of 3422 measurements of soil microbial biomass densities from 14 different biomes. This data set contains different types of methods for measuring microbial biomass, where a large fraction of the data is based on either the fumigation extraction method (112) or the fumigation incubation method (113). Xu et al. assumed the data in their data set represent measurements of the top 30 cm of the soil profile. To extrapolate globally the total biomass of soil microbes in the top 30 cm of the soil profile from the local samples, Xu et al. fit a linear model to predict soil microbial biomass based on globally available environmental parameters such as annual precipitation, annual temperature and soil organic carbon. The specific parameters of the model are described by Xu et al. (110). This extrapolation yielded an estimate for the total biomass of soil microbes in the top 30 cm of the soil profile. To extend the estimate from the top 30 cm to the top 1 meter of the soil profile, Xu et al. assumed the distribution of microbial biomass with depth follows the biomass density versus depth profile of root biomass (21). This assumption was supported by fitting a subset of the data in 5 biomes, for which depth dependence was available, against the distribution of root biomass in the respective biome. Xu et al. used a biome specific vertical root distribution to extrapolate the total microbial biomass in the top 1 meter of soils from the estimates of the total microbial biomass in the top 30 cm in each biome. The specific parameters for this extrapolation are described in Xu et al. In total, Xu et al. estimate the global biomass of soil microbes in the top 1 meter to be ≈23 Gt C.

The second paper by Serna-Chavez et al. used 414 samples of microbial biomass density, measured using the fumigation extraction method (112). From those local samples, Serna-Chavez et al. extrapolated the global biomass of soil microbes by using a multivariate model explaining ≈50% of the variance in microbial biomass density across samples. The model uses environmental parameters such as moisture availability, mean annual precipitation, evaporation and temperature, and soil pH and nitrogen to predict the microbial biomass at each location. Serna-Chavez et al. used global climate maps to derive the environmental parameters globally, and then used these parameters in combination with the model to predict soil microbial biomass globally. The specific parameters of the multivariate model are detailed in (111). Serna-Chavez et al. estimate a total biomass of soil microbes of ≈15 Gt C.

To generate our best estimate for the total biomass of soil microbes, we use the geometric mean of the total estimates from Xu et al. and Serna-Chavez et al. which is ≈18 Gt C (link to full calculation). Both Xu el al. and Serna-Chavez et al. only consider microbial biomass in the top 1 meter of soil. If indeed microbial biomass follows the trend of root biomass reported in Jackson et al., ≈98% of the total microbial biomass is found in the top 1 meter of soil (link to full calculation). Nevertheless, in the estimates made by Whitman, bacterial cells below 1-meter depth constitute ≈15-≈30% of the total amount of bacterial cells in soils. To take into account possible contribution of microbial biomass from layers deeper than 1 meter, we calculate the geometric mean of the fractions of biomass of microbes reported in Jackson et al. and Whitman et al. We thus add ≈7% to our estimate for the total biomass, and our best estimate for the total biomass of soil microbes is ≈20 Gt C (link to full calculation).

To infer the fungal portion of the soil microbial biomass, we used estimates for the fraction of fungi out of the total microbial biomass (114). Joergensen & Wichern (114) estimated the fraction of fungal biomass out of the total soil microbial biomass using several independent methods. These methods include direct microscopic counts, and measurements of cell wall components which are characteristic to either fungi or bacteria (114). Joergensen & Wichern report the fraction of fungal biomass out of the total biomass of soil microbes from different studies conducted in four different soil types (arable, forest, grassland and litter), and using the above-mentioned two different methods. To generate our best estimate for the fraction of fungal biomass out of the total biomass of soil microbes, we first calculated the geometric mean of all the studies conducted in a specific soil type and using a specific measurement method. Taking the geometric mean of values from different soil types but using the same method, we then calculated a characteristic value for each method. For our final best estimate for the fraction of fungal biomass out of the total biomass of soil microbes, we use the geometric mean of the characteristic values from the two methods. The geometric mean of the fraction of fungal biomass out of the total biomass of soil microbes is ≈60% (link to full calculation). This translates to an estimate of ≈12 Gt C for the biomass of soil fungi (link to full calculation). This value is consistent with independent reports of the biomass of bacteria and fungi in the soil (115).

As a consistency check for the estimate, we compare our results to other reports which measured the biomass density of soil fungi locally. A recent study (116) used several independent approaches and arrived at an estimate of ≈0.22-0.68 mg dry fungal mass per gram of soil in a

coniferous forest. This translates to ≈0.1-0.3 mg C per gram of soil. From the dataset in Xu et al. (110), coniferous forest soils contain ≈40 mmol C per kg soil of microbial biomass, which translate to ≈0.5 mg C microbial biomass per gram of soil. Thus, the mass fraction of fungi out of the total microbial biomass is estimated to be ≈0.2-0.6, in line with the results from Joergensen & Wichern (114).

The estimate of 0.1-0.3 mg C per gram of soil also fits well with an analysis from experts in the field, which estimate ≈100 meters of fungal hyphae per gram of soil (117). Using a characteristic value for the diameter of fungal hyphae of ≈3 μm (118), 100 meters of fungal hyphae translates into a biovolume of $\approx 3 \times 10^9$ μm^3 per gram. Assuming tissue density of ≈1 g cm^{-3}, 70% water content and 50% carbon content of dry mass, we arrive at about 0.5 mg C per gram of soil, in the range of the estimates by Baldrian et al. (116).

We now analyze the associated uncertainty of the estimate for the total biomass of soil fungi, which we report as a fold change factor from the mean representing a range akin to a 95% confidence interval of the estimate. As noted above, our estimate of the total biomass of fungi is generated by multiplying our best estimate for the total biomass of soil fungi by our best estimate for the fraction of fungi out of the total biomass of soil microbes. To generate our projection for the uncertainty of the biomass of fungi, we combine the uncertainties associated with both parameters. The estimate of the biomass of soil microbes is based on the geometric mean of the estimates made by Xu et al. (110) and Serna-Chavez et al. (111). Xu et al. do not report uncertainty ranges for the estimate of total biomass but do report 95% confidence intervals for the average microbial biomass densities in each biome. These ranges are all less than 1.5-fold from the average value. Serna-Chavez et al. (111) report 95% confidence interval for the total biomass estimate. This confidence interval is 0.007 Gt C, which is extremely small relative to the total estimate made by Serna-Chavez et al. of 14.6 Gt C. The origin of this confidence interval is not explicitly specified by Serna-Chavez et al., but most probably represents the 95% range of values generated by bootstrapping the parameters of the model which predicts the biomass density of soil microbes. The inter-study variability in the biomass of soil microbes can be calculated by comparing the total estimates provided by Xu et al. and Serna-Chavez et al. The 95% confidence interval for the inter-study variability of the total estimates is ≈1.6-fold (link to full calculation). As this confidence interval is larger than the uncertainties reported in either Xu et al. or Serna-Chavez et al., we use it as the basis for our estimate of the uncertainty associated with the biomass of soil microbes. We also consider the impact of the uncertainty in the fraction of biomass of soil microbes in soil layers deeper than 1 meter on the uncertainty of the total biomass of soil microbes. We calculate the intra-study uncertainty in the values for the fraction of biomass in soil layers deeper than 1 meter reported in Jackson et al. and Whitman et al. as well as the uncertainty of the mean fraction between the estimates of Jackson et al. and Whitman et al. Overall, the uncertainty of the fraction of biomass of microbes in soil layers deeper than 1 meter results in an ≈1.3-fold uncertainty in the biomass of soil microbes (link to full calculation). Combining the uncertainty of the total biomass of soil microbes in the top meter and the uncertainty of the fraction of biomass of soil microbes in soil layers deeper than 1 meter, we estimate a total uncertainty of ≈1.7-fold for the biomass of soil microbes (link to full calculation).

There are several sources of uncertainty that are hard to quantify but increase the total uncertainty associated with the total biomass of soil microbes. First, the two studies reporting the total biomass of soil microbes are not completely independent in terms of the data sets and methods for measuring microbial biomass they use, and thus the interstudy variability does not represent variability between two completely independent studies. Second, the estimates provided in Xu et al. and Serna-Chavez et al. are based on models predicting the biomass density of soil microbes. The models explain only a part of variability in the locally measured biomass density of soil microbes. Remaining sources of variability not explained by the model introduce additional uncertainty to the estimates based on the model. Additionally, both in Xu et al. and in Serna-Chavez et al., the total biomass of soil microbes is extrapolated to a depth of 1 meter by using different empirical equations. It is not clear what is the uncertainty associated with using such equations to extrapolate the biomass of soil microbes to deeper soil layers. Finally, the uncertainty regarding possible contribution of microbes from layers deeper than 1 meter is also hard to quantify. Thus, to take into consideration these extra possible sources of uncertainty, even if not in full, we estimate an uncertainty of ≈2-fold associated with the total biomass of soil microbes (link to full calculation).

The estimate of the fraction of fungal biomass out of the total biomass of soil microbes is based on combining two independent methods, used to estimate the biomass fraction of fungi in four different soil types. The maximal 95% confidence interval of the mean fraction reported from different studies in the same soil type and using the same method is about 3-fold (link to full calculation). The 95% confidence interval of the biomass fraction of fungi averaged across different soil types using the same method is about 1.6-fold (link to full calculation). The 95% confidence interval of the biomass fraction of fungi averaged across different measurement methods is ≈2-fold (link to full calculation). Therefore, we chose to use the largest uncertainty and estimate ≈3-fold uncertainty associated with the fraction of fungal biomass out of the total biomass of soil microbes. Combining these two uncertainties leads us to project that the uncertainty of the estimate of the biomass of soil fungi is about 3-fold (link to full calculation).

Next, we set out to estimate the biomass of mycorrhizal fungi, which are symbionts of plants. Mycorrhiza are generally classified into two main types, arbuscular mycorrhiza, and ectomycorrhiza. Arbuscular mycorrhiza are fungal symbionts that penetrate the cell wall of their host cells. Plants associated with this type of fungi predominate in deserts, grasslands, shrublands, and tropical forest ecosystems (119). Ectomycorrhiza are also fungal plant symbionts, but they do not penetrate the cell walls of their hosts. They are found mainly in boreal and many temperate forests (for example, those dominated by *Pinus*; 120).

Ectomycorrhiza

For the calculation of ectomycorrhizal biomass, we rely on two recent studies that estimated the ectomycorrhizal biomass of Norway spruce dominated forests (121, 122). Both studies use estimates of the total fine-root biomass densities (most ectomycorrhizal fungi are located on fine-roots) in boreal forests, in conjunction with the fraction of biomass that ectomycorrhizal fungi contribute to fine-root biomass. Stögmann et al. (121) estimated the fraction of biomass contributed by ectomycorrhizal fungi to be ≈30%, based on microscopy and 3D reconstruction of

root tips. This value is in line with the reported ranges in the literature. The study by Stögmann et al. (121) estimates ≈18 g C m^{-2} (assuming 50% carbon content). This estimate takes into account only the top 5 cm of soil, and Stögmann et al. (121) state that the fungal biomass could increase significantly if deeper soil layers are included. Comparing the fine-root biomass measured by Stögmann et al. (121) to published censuses of fine-root biomass in boreal forests (21, 123), the values by Stögmann are about 6-fold larger than the characteristic fine-root biomass densities reported. Other studies have estimated the ectomycorrhiza biomass in an Estonian Norway spruce forest at a much lower ≈3 g C m^{-2} (122), even while extending to deeper soil depth. Ostonen et al. (122) rely on a lower biomass fraction of fungi of ≈18%, with a fine-root biomass density estimate that is comparable to values reported in the literature. To estimate the total biomass density of ectomycorrhizal fungi, we correct the ectomycorrhizal fungi biomass density estimates by Stögmann et al. (121) and Ostonen et al. (122) for the characteristic fine-root biomass densities, and for the depth profile of root biomass from Jackson et al. (124). We get that the corrected ectomycorrhizal fungi biomass density from Stögmann et al. (121) is ≈12 g C m^{-2} (the original 18 g C m^{-2} corrected for 6-fold less fine-root biomass and for the fact that 75% of the root biomass is located below 5 cm). Stögmann et al. (121) and Ostonen et al. (122) measured ectomycorrhizal fungi biomass down to 40 cm depth, so correcting for the fine-root biomass depth profile, we get a total of ≈6 g C m^{-2}. As the biomass densities of fine-roots in Stögmann et al. (121) and Ostonen et al. (122) are close to the characteristic densities in the literature, no correction for the fine-root biomass densities was needed. Assuming that these values are characteristic of boreal and temperate forests (values in tropical forests are known to be much lower), we multiply it by the boreal and temperate global forest area of ≈25×10^{12} m^2 (23), and get an estimate of ≈0.15-0.3 Gt C. We thus estimate that the total ectomycorrhizal fungi biomass is roughly ≈0.2 Gt C. We note that this estimate is only a rough estimate of the global ectomycorrhiza biomass, as it is based on very local sampling to extrapolate to other boreal and temperate forests. Nevertheless, fine-root densities used for the estimate are the best available representatives of boreal forests, and the biomass fraction of ectomycorrhizal fungi out of fine-roots seem to remain between ≈20%-40% in different tree species (121). In calculating the ectomycorrhiza biomass, we do not take into account extramatrical mycelia - the collection of filamentous fungal hyphae emanating from ectomycorrhizas. As these mycelia are emanating from the roots of plants, they have a higher probability of being accounted for in the independent estimate of soil microbial biomass.

The main parameters affecting the estimate of the biomass of ectomycorrhizal fungi are the biomass density of fine-roots and the mass fraction of ectomycorrhizal fungi out of the fine-root biomass. The uncertainty in the biomass fraction of ectomycorrhizal fungi between different tree species is around ≈1.3-fold (125). The estimate of the fine-root biomass density has a typical uncertainty of ≈1.3-fold (124). Therefore, the combined uncertainty of the estimate is better than 2-fold. When using the estimates from both papers (as updated above), the two values we get for the biomass density of ectomycorrhiza is closer than 2-fold of each other, which suggests less than 2-fold uncertainty. While these are the reported elements of uncertainty in the available papers we could find, we surmise that systematic biases can arise from the very limited geographic sampling performed to date, and we thus evaluate the uncertainty to rather be better than an order of magnitude.

Arbuscular mycorrhiza

For the calculation of arbuscular mycorrhiza biomass, we use a study which estimates the global arbuscular mycorrhiza biomass at 0.7 Gt C (assuming 50% carbon content; 126). The estimate is based on using previous estimates of fine-root length per unit area in different biomes (21). Treseder & Alison (126) total fine-root length to volume by using estimates of the diameter of fine-roots in different biomes. From the total volume of fine-roots, Treseder & Alison (126) use an estimate of the fraction of the length of the fine-root covered with arbuscular mycorrhizal fungi. Arbuscular mycorrhizal fungi constitute only a fraction of the volume of the fine-root, and Treseder & Alison (126) use an estimate of this fraction to convert the total volume of arbuscular mycorrhizal fungi covered fine-root to the volume of the fungi alone. Treseder & Alison (126) then convert this volume to biomass by using estimates of the fungal tissue density. Treseder & Alison (126) only considered fungi from the 10-top cm of the soil. To account for biomass in lower soil layers, we used the depth profile of roots in different biomes (21) to calculate the fraction of total root biomass that reside in the top 10 cm. We then corrected our biomass estimates to account for these extra roots in lower depths, which brings the total estimate from 0.7 Gt C to ≈2 Gt C. This is probably an overestimate of the biomass of arbuscular mycorrhiza as the fraction of roots covered with fungi decreases with depth (127). As we discuss above, the main source we base our estimate of the biomass of arbuscular mycorrhizal fungi upon is the study by Treseder & Alison (126). This was recommended by experts in the field for estimation of arbuscular mycorrhizal fungi abundance (117). Correcting the original estimate of arbuscular mycorrhizal fungi made by Treseder & Alison (126) for the dry mass density used, as well as for the characteristic radii of fine-roots, we estimate about ≈0.15 Gt C.

In establishing the uncertainty, we focus on several parameters that affect the estimate of the biomass of arbuscular fungi biomass. The first is the total length of fine-roots per unit area. This parameter has an associated uncertainty of ≈1.3-fold (124). A second parameter affecting the estimate is the fraction of fine-root length covered by arbuscular mycorrhizal fungi. This fraction has an uncertainty of ≈1.6-fold (126). The diameter of the fine-roots will influence the estimated volume of fine-roots and has an uncertainty of ≈1.2-fold (124). The fraction of the volume of fine-root colonized by arbuscular mycorrhizal fungi will also affect the final estimate and has an uncertainty of ≈1.1-fold. Finally, the tissue density will affect the conversion between fungal volume and biomass and has an uncertainty of ≈1.5-fold. Overall, we project the uncertainty of the estimate of biomass of arbuscular mycorrhizal fungi to be about 2-fold.

Marine fungi

To estimate the total biomass of marine fungi, we consider several environments in which fungi reside: planktonic fungi in the epipelagic layer of the ocean (between 0-200 meters in depth), the planktonic fungi in the deep-sea (below 200 meters) and particle-attached fungi. For planktonic fungi in the epipelagic layer, the DNA copy number of planktonic fungi was measured using qPCR to be about ≈10% of the prokaryotic DNA copy number at the Pacific Warm Pool (128). As an independent method for estimating the ratio between the biomass of fungi and prokaryotes, we rely on direct counts of fungal cells and prokaryotes, along with their biovolumes, at the

coastal upwelling ecosystem off central Chile (129) The mean ratio between the biomass of planktonic fungi and prokaryotes was ≈25% (link to full calculation). As our best estimate for the total biomass of marine fungi in the epipelagic layer we use the geometric mean of the estimates based on qPCR and direct counts, which is ≈20%. As we estimate ≈0.4 Gt C of epipelagic prokaryotes, this translates to ≈0.07 Gt C if extrapolated on a global scale. This is inline with recent metabarcoding survey of eukaryotic diversity in the sunlit ocean has not identified fungi as a major constituent of the plankton community (130). In the deep ocean, recent studies have identified fungi as contributing ≈15% of the total 18S rDNA of microbial eukaryotes (131). However, the total biomass of these deep-sea microbial eukaryotes is estimated at ≈0.1 Gt C (see marine protists section; 132), so we estimate ≈0.015 Gt C of deep-sea planktonic fungi. However, these studies were mainly focused on measuring the biomass of heterotrophic protists, and thus they might capture only unicellular fungi and not filamentous fungi. To take into account the possibility of deep-sea filamentous fungi, we extend our estimate of the ratio between planktonic fungi and prokaryotes to the mesopelagic and bathypelagic realms, which results in an estimate of a total of ≈0.17 Gt C of deep-sea fungi (link to full calculation). As our best estimate of the biomass of planktonic deep-sea fungi, we use the geometric mean of the estimates we generated based on 18S rDNA sequencing and based on applying the ratio of fungal and prokaryotic biomass from the epipelagic layer to the mesopelagic and bathypelagic layers. Thus, our best estimate of the biomass of deep-sea planktonic fungi is ≈0.05 Gt C.

For particle-attached fungi, a recent study by Bochdansky et al. (50) has measured the relative biomass of particle-attached fungi and prokaryotes in the bathypelagic layer. We use the values reported in Bochdansky et al. to estimate that the biomass of particle-attached fungi is about 70% of particle-attached prokaryotes in the bathypelagic layer. We could not find estimates for the ratio between the biomass of particle-attached fungi and prokaryotes in shallower layers of the ocean, and thus we apply the ratio measured in the bathypelagic layer across the entire volume of the ocean. We arrive at an estimate ≈0.3 Gt C of particle-attached prokaryotes, so we estimate ≈0.2 Gt C of particle-attached fungi (link to full calculation). Overall, we estimate that the biomass of marine fungi is ≈0.3 Gt C (link to full calculation), which is small compared to our estimate of ≈12 Gt C of soil fungi. A general caveat associated with estimating the biomass of marine fungi is that this group has been understudied due to accessibility challenges, which can lead to underestimates in their biomass contribution (for example due to non-optimal sampling techniques). As the data regarding marine fungi is scarce, we chose to project an uncertainty of an order of magnitude for our estimate of the total biomass of marine fungi.

Deep subsurface fungi

Another contribution to the global biomass of fungi comes from terrestrial and marine deep subsurface environments. The biomass of fungi in both environments has not been extensively studied. For the marine deep subsurface, biomass appears to be very low, as few fungal cells have been observed in subsurface samples (133). In some sites, fungi account for 3-20% of the total metatranscriptomic reads (134), but there is no strongly supported quantitative connection between the fraction of total reads represented by fungi and their relative biomass. Ribosomal DNA (rDNA) copies of eukaryotes appear to be about 2 orders of magnitude lower than those of

prokaryotes in the marine deep subsurface (135), even though rDNA copy number may not predict biomass reliably. For the terrestrial deep subsurface, only a handful of measurements have been conducted, using either culture-based or culture-independent methods (136, 137). These studies have suggested that the abundance of fungal cells is five to six orders of magnitude lower than the abundance of prokaryotes. Even when accounting for the larger carbon content of fungal cells, the biomass of fungi based on these measurements will still be much lower than that of prokaryotes. Other studies based on hybridization of oligonucleotide probes to rRNA extracted for aquifers have suggested eukaryotic sequences to account for ≈10% of the total rRNA. It is not clear, however, if this rRNA is fungal in origin, and if this type of measurement is quantitatively correlated to the relative biomass of eukaryotes.

Recent sequencing-based studies report diverse communities of active fungi in the terrestrial deep subsurface (138), but these studies do not report absolute densities of fungal biomass in these environments. Fungi might also be present in the oceanic crust, but at present we could not find reliable sampling of their abundance in this environment. Due to the extreme scarcity of data on the abundance of fungi in the terrestrial deep subsurface and oceanic crust, we do not include them in the current analysis. We highlight this knowledge gap in the discussion section of the paper main text.

Total fungal biomass

Summing the estimates for the biomass of soil fungi, ectomycorrhizal fungi, arbuscular fungi, and marine and deep subsurface fungi we get a total of ≈12 Gt C (link to full calculation). One caveat of combining these estimates together is the risk of double counting mycorrhizal biomass. The soil microbial biomass is based on bulk soil samples. These bulk samples contain in them mycorrhiza from nearby plants. Therefore, some of the mycorrhizal biomass will be counted twice in the soil fungi section and the relevant mycorrhiza section. Because the global biomass of ectomycorrhiza and arbuscular mycorrhiza are small relative to the total soil fungal biomass, even a hypothetical double counting would not affect our results significantly. In total, we estimate **≈12 Gt C** of fungi.

Active and inactive microbial biomass

An important point to take into account when estimating the total biomass of microbes is their high metabolic flexibility. Microbial cells can be found in various physiological states. We are used to thinking of cells as a biological entity which is dividing vegetatively, with a range of growth rates governed by the environmental conditions. Under starvation conditions, microbes can also sustain their viability without increasing their biomass. These types of cells do not have an associated growth rate but are turning over their biomass (and hence carbon) at a very slow rate (139). An alternative strategy to cope with harsh environmental conditions is to enter a physiological state of dormancy, in which the metabolic activity is further extremely reduced. Dormancy usually requires cells to invest resources into resting structures (such as spores) and the machinery that is needed for transitioning into and out of a dormant state (140). Finally, cells can also be in the process of cell death, with a varying amount of cellular structures intact (ribosomes, cell membrane etc.).

Different techniques used to measure the biomass of microbes might include or exclude different types of physiological states from the tally. In theory, we would like to include in our account all physiological states except dead cells. In most of the studies we base our estimates upon, there is no explicit discussion of the type of physiological states which the assay accounts for. We proceed to characterize to the best of our knowledge the type of physiological states included in our estimates in each environment.

A common practice to distinguish between cells which are alive or dead is to compare the number of cells stained with a DNA dye, which represents the total number of cells, with the number of cells that are stained with a fluorescent probe for 16S rRNA (using FISH or similar methods), which represents the population of cells containing ribosomes. Based on this type of assay, people have characterized the number of cells in a dormant state out of the total population (140). A recent meta-analysis by Lloyd et al. (28) has characterized the fraction of cells containing ribosomes out of the total population of cells in the marine environment and in subseafloor sediments. Lloyd et al. found that in the marine environment around ≈80% of bacteria stained for DNA also stained with 16S rRNA probes. As our estimates for the biomass of marine bacteria and archaea is based on counting cell numbers after DNA stains, this means that the actual number of living cells might be about 20% lower than our estimate.

For the marine deep subsurface environment, Lloyd et al. found that the fraction of cells stained with 16S rRNA probes is highly dependent on the technical parameters of the protocol. In studies which use best practice protocol parameters (using CARD-FISH and permeabilizing cells with proteinase K), the fraction of cells stained with a 16S rRNA probe was ≈84%. These values are in line with an independent study which measured the fraction of cells in subseafloor sediments which are able to assimilate labeled carbon and nitrogen compounds (141). This means that our estimates for the total number of cells in the marine deep subsurface, which are based on DNA stains, might similarly overestimate the actual number of cells by about 20%. Another lesson from the meta-analysis by Lloyd et al. is that measuring a low fraction of cells which are stained by 16S rRNA probes does not necessarily imply that a large fraction of the cells is dormant or dead but could also imply that there might be a technical issue with the parameters used for the assay. In other environments, such as soils and the terrestrial deep subsurface, no comprehensive studies were conducted. A study using CARD-FISH in different soils has found the fraction of cells stained with a 16S rRNA probe ranged between 51-94% of the total cell count, similar to values in the marine environment (70). In order to have a fair comparison between the different environments, we chose to report estimates based on the total number of cells stained by a DNA stain.

A physiological state which is likely to be underrepresented in our data is spores. Specific types of spores, such as endospores, are unlikely to be stained by fluorescent DNA dyes or by 16S rRNA staining techniques (142). In addition, spores could also be insensitive to the chloroform fumigation extraction method, which is used to estimate the biomass of soil microbes (143). In some environments, such as in subseafloor sediments, endospores might contribute a significant fraction of biomass, similar to that of non-spore cells (142). However, due to scarcity of data regarding the abundance of spores in all environments, we could not comprehensively include

spores as part of our analysis. This is not to say that spores are completely unrepresented, as there are many types of spores, some of which might be amenable to the measurement techniques on which we based our estimates for the total biomass of microbes. For the same reason we could not include in this analysis encysted forms of protists, which are akin to the spores of bacteria and archaea. We hope further research will clarify the relative biomass contribution from the different morphological states of microbes.

Annelids

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Annelids are a phylum of animals containing both marine and terrestrial species. For estimating the terrestrial annelid biomass, we followed a previous study (144) that chose as the dominant contributions the terrestrial earthworms and the worm family Enchytraeidae. We base our estimate on their values for the density of annelid biomass in the soil (144). The estimate in Fierer et al. (144) is based on combining data for different locations. For some locations there are direct biomass measurements and for other locations biomass is estimated based on conversion from number of individuals to biomass (based on size or mass) to give the biomass per unit area in each biome. We multiplied these biomass values in Fierer et al. (144) by the land area of each biome to reach the final estimate. This estimate does not include earthworm biomass in leaf litter (see litter section below). The data also does not contain depth data, but since most soil faunal biomass is likely to be restricted to the top 15 cm of soil (145), this is unlikely to lead to a significant underestimation of faunal biomass in most soils, although desert soils may be an exception. Fierer et al. (144), give the mean biomass densities for each biome in g C m^{-2}, as well as the median biomass densities. To estimate the global biomass of annelids, we calculate a total estimate based independently on the mean and then separately on the median biomass densities. In each estimate we multiply the mean (or median) biomass density for each biome by the total area of that biome (link to full calculation). Our best estimate is the geometric mean of the estimate based on mean biomass densities and the estimate based on median biomass densities. Two important biomes are missing from the analysis made by Fierer et al. (144), namely tropical grasslands (savannas) and crops. We compiled estimates for the biomass density in each of the missing biomes by performing a literature survey (146–150). In tropical grasslands, two unique environments are present, each with its own characteristic biomass density. The first is natural savanna, and the second is pasture. As we do not know the division of area between the area covered by natural savannas and pastures, we tested the sensitivity of our final estimate to this parameter. The final estimate varied between ≈0.05 Gt C and ≈0.4 Gt C. Therefore, we chose as our best estimate **≈0.2 Gt C** (link to full calculation). It is worth noting that estimates for different biomes have different sampling sizes and therefore different uncertainties associated with them. We are aware that biomass densities with low certainty may introduce biases into the global estimate. As a sanity check on this result, a recent paper (117) has produced approximate estimates for the abundance of various taxa in the soil by consulting with experts for each taxon. The study reports an estimate of 300 individual earthworms per m^2. Combining this estimate with the average biomass of an individual earthworm used in Fierer et al. (144) which is 5 mg C, we

get ≈1.5 g C m^{-2}. This value lies in the range of reported biomass densities in (144) of 0.2-5 g C m^{-2}, thus strengthening our confidence in the estimate.

Annelids are also present in marine environments. Polychaetes are present in both pelagic and benthic environments. Pelagic polychaetes are part of the macrozooplankton. Analysis of macrozooplankton biomass is detailed in the marine arthropod section. The analysis shows that the main fractions of macrozooplankton biomass are contributed by cnidarians, molluscs and arthropods. Therefore, the biomass contribution from pelagic annelids will probably not dramatically affect our estimate. We could not find any study on the contribution of annelids to the total benthic biomass, and therefore they are not added here. For an upper limit and further discussion of benthic biomass see the "Other phyla" section.

Nematodes

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Nematodes are also an animal phylum containing both marine and terrestrial species. To estimate terrestrial nematode biomass, we use the same study as for estimating annelid biomass (144). As stated in the annelid section, this study estimates the global distribution of nematode biomass in the soil. The estimate in Fierer et al. (144) is based on combining data for different location. For some locations, there are direct biomass measurements and for other locations biomass is estimated based on conversion from number of individuals to biomass (based on size or mass) to give the biomass per unit area in each biome. We multiplied the values in Fierer et al. (144) by the land area of each biome to reach the final estimate. This estimate doesn't include nematode biomass in leaf litter (see litter section below). The data also doesn't contain depth data, but since most soil faunal biomass is likely to be restricted to the top 15 cm on the soil profile (145), this is unlikely to lead to a significant underestimation of faunal biomass in most soils, although desert soils may be an exception. Fierer et al. (144), give the mean biomass densities for each biome in g C m^{-2}, as well as the median biomass densities. To estimate the global biomass of nematodes, we calculate a total estimate based separately on the mean and median biomass densities by multiplying the mean (or median) biomass density for each biome by the total area of that biome. We use as best estimate the geometric mean of the estimate based on mean biomass densities and the estimate based on median biomass densities. Two important biomes are missing from the analysis made by Fierer et al. (144), namely tropical grasslands (savannas) and crops. We could not find reliable sources for the biomass densities of nematodes in those environments, and thus we chose to extrapolate the biomass density we got for the other biomes. We get an estimate of ≈0.006 Gt C (link to full calculation). It is worth noting that estimates for different biomes have different sampling sizes and therefore different certainty associated with them. We are aware that biomass densities with low certainty may introduce biases into the global estimate. To compare this estimate with the contribution of marine nematodes, we use a study which estimates the total seafloor biomass, and the contribution of different classes of organisms to the total biomass (87). Wei et al. (87) estimate a global biomass of 0.1 Gt C for the seafloor, with ≈13% contributed by meiofauna, which includes, among other organisms, nematodes. Nematodes are a major

constituent of the meiofauna (151, 152), so we estimate marine benthic nematode biomass at ≈0.01 Gt C (link to full calculation). This estimate does not take into account nematode biomass in environments such as seamounts and submarine canyons, because there is very little knowledge of the biomass of nematodes in those locations. An upper limit and further discussion of benthic biomass is given in the "Other phyla" section. Nematodes have also been found in the terrestrial deep subsurface (153), however their population densities are significantly lower than those of prokaryotes in the terrestrial deep subsurface. For every nematode, there are 10^{10}-10^{12} prokaryotes (153). Nematodes in the terrestrial subsurface have a carbon content of ≈10^{-8} g C (153), which is ~10^{6} times larger than prokaryotes. Thus, the biomass of terrestrial deep subsurface prokaryotes is roughly 10^{5} times that of nematodes. We thus estimate that the total biomass of nematodes in the terrestrial subsurface is ~0.001 Gt C, which is negligible compared to the biomass of soil and benthic nematodes. In summary, the global nematode biomass is estimated at **≈0.02 Gt C**. As a sanity check on this result, a recent paper (117) has produced approximate estimates for the abundance of various taxa in the soil, consulting with experts for each taxon. They estimate 10^{5}-10^{7} individual nematodes per m^2. Combining this estimate with the average biomass of an individual earthworm used in Fierer et al. (144), which is 0.05 µg C, we get ≈0.05 g C m^{-2}. This value lies in the range of reported biomass densities in Fierer et al. (144) of 0.01-0.3 g C m^{-2}, thus strengthening our confidence in the estimate.

Litter microbes and fauna

Our estimates for the biomass of soil microbes and fauna do not take into account the biomass of the respective taxa in litter. In this section we give a crude estimate of the total biomass of litter microbes and fauna (prokaryotes, protists, fungi, nematode, annelids, and arthropods). Most studies report the abundance of litter flora and fauna per unit mass of litter. In order to estimate the total abundance of the litter flora and fauna, we need to know the total mass of litter. Estimates based on direct field observations put the global pool of dry weight from plant litter at about ≈80 Gt C (7, 154). For microbes (bacteria, archaea, fungi and protists), Recent studies estimate the microbial biomass in litter to be less than 1% of the total organic carbon mass in litter (155–158). This means that the amount of microbial biomass in litter will not surpass 1 Gt C, which is less than 10% of our estimate for the total biomass of soil microbes (see section on soil fungi and soil bacteria and archaea). Thus, the contribution from litter microbes will likely not affect dramatically our estimates for the total biomass of soil prokaryotes, protists or fungi.

Our estimates of the total biomass of annelids and nematodes is based on data collected in Fierer et al. (144). Fierer et al. state that many of the data on the biomass of soil fauna may contain measurements of both the soil and litter communities. This means that our estimate includes, even if to a small degree, also the biomass contribution of litter nematodes and annelids. We argue that the biomass contribution from litter nematodes and annelids not accounted for by our estimates will likely not change the final estimates. To back this claim, we use the total mass of litter worldwide. If we divide the total biomass of the litter pool by the total ice-free terrestrial area, which is ≈1.3×10^{14}, we arrive at an estimate of ≈0.5 kg m^{-2}. Most of the biomass of nematodes and annelids is found in the top ≈10-20 cm of soil (144), which weighs about ≈200 kg m^{-2}, assuming soil bulk density of ≈1.3 g cm^3. This means that the total weight of soil is ≈400-fold

larger than the weight of litter (litter weighs ≈0.5 kg m^{-2}, and the relevant soil section weighs ≈200 kg m^{-2}) and thus in order to have a significant contribution of biomass from litter, abundances in litter should be about two orders of magnitude larger than in soil. This does not seem to be the case for nematodes (159), as well as for annelids (160). For arthropods, we already incorporated the biomass of litter arthropods in our estimates.

Chordates

Fish

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Global fish biomass was historically estimated based on trawling experiments and primary productivity coupled to models of trophic transfer efficiencies (161–163). These estimates put global fish biomass at around ≈0.3 Gt C (2 Gt fresh weight), with half of fish biomass contributed by mesopelagic fish (fish that live at 200-1000 m depth; 161), which corresponds well with trawling based estimates of mesopelagic fish biomass (163). In 2010, an expedition named the Malaspina explored, among other goals, the biodiversity of the deep ocean. As part of this effort, a recent study (164) has used acoustic observations from the Malaspina campaign to estimate the global mesopelagic biomass. Mesopelagic fish reflect sonar signals and they are a strong feature visible all across the oceans. An estimate of the mesopelagic biomass is based on measurement of acoustic backscatter from the deep scattering layer, combined with calibrations of for the strength of reflection of acoustic signals from a single mesopelagic fish, termed target strength. The study by Irigoien et al. (164) measures the total scattering along the course of the campaign and finds a correlation between this scattering strength and the local net primary productivity (NPP). Then, using estimates of NPP across the ocean, Irigoien et al. (164) extrapolates the total scatter across the entire ocean. The specific parameters of the correlation between NPP and scattering strength are given in detail in (164). Irigoien et al. convert this total scatter to biomass by using relations of the target strength per unit biomass of a single fish. The conclusion of the analysis by Irigoien et al. puts mesopelagic biomass at ≈1.5 Gt C, dominating all other fish populations. Independent estimates of the total backscatter from the deep scattering layer (165) are about ≈40% lower than the estimate by Irigoien et al. (164). We thus update the original estimate by Irigoin et al. by using the geometric mean of the total backscatter estimated by Irigoien et al. and Proud et al. (165). To convert the total backscatter to fish biomass, we relied on the data in Irigoien et al. for the target strength of mesopelagic fish. The measurements of target strength can be divided to two main categories, fish with and without a swim bladder. We use the mean target strength for each group (with and without swim bladder). We assume the population of mesopelagic fish is divided equally (for lack of better data) between fish with and without swim bladder and calculate a mean population target strength. Combining the total backscatter and the mean target strength we estimate a total biomass of ≈1.8 Gt C for mesopelagic fish (link to full calculation). As we discuss below, estimating biomass based on acoustic scattering has several caveats which might bias the results. The mesopelagic fish community is highly diverse, and thus using a "characteristic" target strength per unit biomass inserts significant uncertainty to the final estimate (166).

Nevertheless, previous estimates based on trawling are known to underestimate mesopelagic fish biomass due to escapes and avoidance of trawls (167). While the recent estimate by Irigoien et al. (164) is highly divergent from previous estimates, we rely as in other cases on the geometric mean to give a robust estimator of global biomass of fish. Previous trawling-based estimates put mesopelagic fish biomass at ≈0.15 Gt C (163). The geometric mean of the two estimates is ≈0.5 Gt C (link to full calculation). We add to this geometric mean the contribution from other group of fish of ≈0.15 Gt C (161), to yield a final estimate of **≈0.7 Gt C** (link to full calculation).

We now analyze the associated uncertainty of the estimate for the total biomass of fish, which we report as a fold change factor from the mean, representing a range akin to a 95% confidence interval of the estimate. We start by analyzing the associated uncertainty of the estimate based on acoustic measurements. The main parameters contributing to the uncertainty are the total scatter that is extrapolated globally, and the target strength used to convert this scatter to biomass. The correlation between the local NPP and the local scattering strength, explains ≈60% of the variance in the acoustic scattering. Different kinds of regression between local NPP and scattering coefficients give variability of ≈1.2-fold in the total backscatter of the deep scattering layer (link to full calculation). The inter-study uncertainty associated with the total backscatter is ≈1.7-fold (link to full calculation). The measurements of target strength can be divided to two main categories, fish with and without a swim bladder. The scattering of each group has characteristic uncertainty of ≈50%. We modelled different population compositions between the two groups to project the uncertainty associated with the unknown distribution of mesopelagic fish with and without swim bladder. We assume fish with swim bladder account for at least 20% of the population by mass, which is a reasonable assumption given that both variants appear in most species (164). Under this assumption variation in the relative abundance of fish with and without swim bladder will change the total estimate for the biomass of mesopelagic fish by ≈8-fold (link to full calculation). As we could not find any rigorous analysis of the composition of mesopelagic fish, we use this range as the uncertainty associated with the distribution of fish with and without swim bladder. Combining these sources of uncertainty, we project that the uncertainty associated with acoustics-based estimate for the biomass of mesopelagic fish is ≈10-fold (link to full calculation). Another uncertainty to consider is the fraction of the backscatter resulting from mesopelagic fish biomass rather than other organisms such as crustaceans, and other members of mesopelagic plankton. Previous reports suggest that most of the backscatter is originating from mesopelagic fish (168, 169), however estimating the uncertainty associated with this aspect is difficult. Additional sources of uncertainty, which are also hard to quantify, stem from the population structure of mesopelagic fish. Small fish which are smaller than the wavelength of the ultrasonic sonar signal will have lower backscatter due to Rayleigh scattering, and thus this type of acoustic estimate might underestimate the total biomass. Conversely, because acoustic models of the relation between backscattering and fish length are not completely linear (166), a change in the size spectrum of the mesopelagic fish community might influence the backscatter. Finally, possible effects of resonance, mainly from swim bladders, might affect the estimate, even though Irigoien et al. (164) gives evidence that such an effect is probably small. Our estimate for the global biomass of mesopelagic fish is based on a geometric mean between the acoustics-based estimate by Irigoien et al. (164) and previous estimates (161–163). Previous estimates based on trawling, which put mesopelagic fish biomass at ≈0.15 Gt C, represent an underestimate of the

global biomass of mesopelagic fish. Conversely, higher estimates by Irigoien et al. (164) of ≈5 Gt C (assuming most of the population contains fish without swim bladder, which have low target strength) is probably an overestimate. The uncertainty in the estimate of the total mesopelagic fish biomass between the two independent methods is ≈10-fold (link to full calculation), which is similar to the ≈10-fold uncertainty associated the acoustics-based estimate. We thus project an uncertainty of about one order of magnitude for the estimate of the biomass of mesopelagic fish. For estimating the biomass of non-mesopelagic fish, we rely on estimates by Wilson et al., which does not report an uncertainty range for the biomass of non-mesopelagic fish. A later study (170), gave an estimate for the total biomass of fish with body weight of 1 g to 1000 kg, based on ecological models. Jennings et al. report a median estimate of ≈0.75 Gt C for all fish biomass, which includes both mesopelagic and non-mesopelagic fish biomass. Jennings et al. report a 90% confidence interval of 0.05-4 Gt C for the global biomass of fish. We take this range as representative of the uncertainty of the non-mesopelagic fish biomass estimate. Combining our uncertainty projections for mesopelagic fish biomass and non-mesopelagic fish biomass, we project an uncertainty of about 8-fold associated with the estimate of the global biomass of fish (link to full calculation).

Humans and livestock

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

In order to estimate global livestock biomass, we use data on global stocks of cattle, sheep, goats, pigs, poultry and other livestock species from the FAOStat database (http://faostat3.fao.org/; domain: Production/Live animals). We multiply the total number of individuals for each species with mass estimates of each species from the IPCC (171). For humans, we use the UN estimate of the global population, and a mean mass per person of 50 kg (172). The global biomass of livestock turns out to be **≈0.1 Gt C** (link to full calculation). Out of this global livestock biomass, we estimate the biomass of poultry at ≈**0.005 Gt C** (link to full calculation). For humans, the estimated global biomass is **≈0.06 Gt C** (link to full calculation).

Wild mammals

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Wild mammal biomass contains both land mammals and marine mammals. To estimate the biomass of wild land mammals, we rely on three approaches. The first is based on Smil (173). Smil estimates 0.004 Gt C of wild land mammals based on estimated biomass densities per biome taken from the History Database of the Global Environment (HYDE; 174). Our second resource for estimating the current biomass of wild land mammals is based on Barnosky (175), which reports a biomass of ≈0.008 Gt C for present day megafauna. As an alternative approach, we used data on the mass of individual mammals for each mammal species, the population density of each species, and the study area in which the population density was measured (176). This dataset

included data on ≈350 mammal species. Using multiple regression, a power-law (log-log) functional relation was established between body mass, study area and the total number of individuals measured in the study area. The functional relation established is: $\log_{10}$(number of individuals) = 1.65-0.53×$\log_{10}$(body mass) + 0.73×$\log_{10}$(study area), (R^2= ≈0.5). This relation was used to extrapolate the total number of individuals for ≈3700 mammal species for which range and mass data is available, by using range sizes from IUCN data (http://www.iucnredlist.org/technical-documents/spatial-data#mammals) as a surrogate for study area, and data on body mass (177; there are additional ≈1800 mammal species without mass data, but these usually have small body mass, small ranges and low population densities). From total number of individuals, the total biomass is calculated by multiplying the total number of individuals by the mean body mass. This approach yielded an estimate of ≈0.001 Gt C. We use the geometric mean of the three approaches as our final estimate for wild land mammals. The biomass of wild land mammals is thus estimated at ≈0.003 Gt C (link to full calculation). As our best projection for the uncertainty associated with our estimate of the biomass of wild land mammals, we calculate the 95% confidence interval around the geometric mean between our three estimates, which is ≈4-fold (link to full calculation).

For marine mammals, we use Christensen (178), which estimate a global whale mass of ≈0.004 Gt C (link to full calculation). This estimate by Christensen is based on using a dataset compiled by Kaschner et al. (179), gathered from various resources. The biomass of marine mammals is dominated by whales and seals. Christensen reports a 95% confidence interval of ≈1.4-fold (link to full calculation). As a consistency check, we compared the data for ≈30 whale species which are the main contributors to the global marine mammal biomass with data from the IUCN. The correlation between the data is high (Spearman R^2=0.975), and the total biomass from both methods varies about ≈1.3-fold (link to full calculation). A caveat of our uncertainty analysis for wild marine mammals is that the estimates from the IUCN and Christensen might not be independent of each other. Together, we estimate the wild terrestrial and marine mammal biomass at 0.007 Gt C (link to full calculation). We combine our uncertainty projections for terrestrial and marine mammals, and thus project the uncertainty associated with our estimate of the total biomass of wild mammals to be ≈2-fold (link to full calculation). We note that interestingly, our analysis suggests that the biomass of whales is higher than the total biomass of all wild land mammals combined.

Wild birds

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Bird biomass was calculated based on an estimate for the total number of birds in the world and on estimated average biomass per individual bird. The estimate of total individual birds is based on work by Gaston & Blackburn (180) using several methods to estimate the total number of birds: Gaston & Blackburn extrapolate from abundance densities for birds found in forests; Gaston & Blackburn extrapolate total numbers from estimates of the number of birds in Africa and in the Nearctic by dividing the total number by the area of the continent and applying the resulting density to the total land area on Earth; Gaston & Blackburn use an empirical relation

between area and the number of birds present in it and extrapolate it to the land area on Earth; Finally, Gaston & Blackburn use an allometric scaling of the density of individuals per species with body size. To convert the population density of individuals per species into total population density, Gaston & Blackburn use an estimate for the average number of species coexisting in a square kilometer. All these different methods yield estimates which are similar, and stand at about $2\text{-}4\times10^{11}$ individual birds.The average biomass of an individual bird was estimated based on the relationship between number of individuals and body size in British birds (181). By analyzing the data in Nee et al. (181), we calculate the average bird wet weight of British birds to be ≈80 g per bird. Assuming 70% water content and 50% carbon content out of the dry weight, we calculate ≈0.004 Gt C of birds biomass globally (link to full calculation).

As an alternative approach, we used data on the mass of individual birds for each bird species, the population density of each bird species, and the study area in which the population density was measured (176). This dataset included data on ≈900 bird species. Using multiple regression, a power-law (log-log) functional relation was established between body mass, study area and the total number of individuals measured in the study area. The functional relation established is: $\log_{10}(\text{number of individuals}) = 3.26\text{-}0.3\times\log_{10}(\text{body mass}) + 0.63\times\log_{10}(\text{study area})$, ($R^2$ = ≈0.3). This relation was used to extrapolate the total number of individuals for ≈75% of the total number of bird species, by using breeding range sizes data (176) as a surrogate for study area, and data on body mass (177). From total number of individuals, the total biomass is calculated through multiplying the total number of individuals by the mean body mass. This approach yielded an estimate of ≈0.001 Gt C (link to full calculation).We take the geometric mean of the estimates from both approaches, **≈0.002 Gt C**, to be our final estimate for the global biomass of birds.

Reptiles

Our estimate of the biomass of reptiles is based on data of mean mass of individuals for about 99% of reptile species. Data on the mean (or minimum) population density for ≈500 species of reptiles was used to generated a power-law (log-log) regression between mean mass of an individual reptile for each reptile species and mean (or minimum) population density for the same reptile species. The relation established based on mean population densities was $\log_{10}(\text{population density}) = 4.4\text{-}0.5\times\log_{10}(\text{body mass})$, ($R^2$=0.13). The relation established based on minimum population densities was $\log_{10}(\text{population density}) = 2\text{-}0.6\times\log_{10}(\text{body mass})$, ($R^2$=0.18). These relations were used to estimate the population density of the missing ≈9500 species of reptiles without population density estimates. From the extrapolated population densities, estimates of the global biomass of reptiles were calculated through multiplying the population densities by the range of each species (182), and its mean mass per individual (183). The estimates using mean or minimum population densities are ≈0.5 Gt C and ≈0.001 Gt C, respectively.
There is a pronounced effect of the size of the area in which the densities were sampled on the density of individuals (176). Therefore, an alternative approach was used. In this approach, we used data on the mass of individual birds, their population density, and the study area in which the population density was measured (176). This dataset included data on ≈192 reptile species. Using multiple regression, a power-law (log-log) functional relation was established between body mass, study area and the total number of individuals measured in the study area. The functional relation established is: $\log_{10}(\text{number of individuals}) = 1.3\text{-}0.13\times\log_{10}(\text{body mass}) +$

$0.5 \times \log_{10}$(study area), ($R^2 \approx 0.17$). This relation was used to extrapolate the total number of individuals for all reptile species, by using range sizes data (182) as a surrogate for study area, and data on body mass (183). From total number of individuals the total biomass is calculated by multiplying the total number of individual by the mean body mass. This approach yielded an estimate of 0.00005 Gt C. We use the geometric mean of these three estimates, 0.003 Gt C, as our best estimate for the biomass of reptiles. This type of analysis has many caveats associated with it, and a very high uncertainty (≈2 orders of magnitude). Because of such high uncertainty and because of the fact that the biomass reported here for reptiles is very small, we chose not to report it in the main text. Nevertheless, this is a first attempt at estimating the global biomass of reptiles, and he hope it will motivate further research that will improve upon our work.

Amphibians

After searching the literature for estimates of the global biomass of amphibians, or data that could be used to produce a reliable estimate, we could not find sufficient data to establish a robust estimate. The studies we did encounter (184–192) provide estimates of biomass densities for only a few species in a specific location. We could not determine the fraction of the population of amphibians the measured species represent, as well as the range each species has. This makes it hard to establish a reliable estimate for the total biomass of amphibians. We argue that the contribution from amphibian biomass will likely not change significantly our estimate of the total biomass of animals. To support this argument, we very crude estimate the total biomass of amphibians. We take data on the distribution of body weight in amphibians from a recent study (193), and calculate a characteristic biomass based on the geometric mean to give ≈1 g C per individual amphibian. We naively assume a density of ≈1 amphibian per m^2 over the entire terrestrial surface area, which is likely to be an overestimate of the density of amphibians. Multiplying the characteristic biomass of a single amphibian by the density of ≈1 m^2, we get a biomass density of ≈1 g C m^{-2}. Applying this density over the entire ice-free terrestrial surface area, which is $\approx 10^{14}$ m^2, we estimate a total biomass of ≈0.1 Gt C. This estimate is likely an overestimate of amphibian biomass, but is still negligible relative to the total biomass of animals, which we estimate at ≈2 Gt C. Because this estimate is very crude and has many caveats associated with it, comparison of amphibian biomass relative to contributions from other vertebrates is not possible. We hope future work will result in a more robust estimate.

Arthropods

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

In estimating the global mass of arthropods, we divide our analysis by environment. We give an estimate for the biomass of terrestrial arthropods, and a second estimate for the biomass of marine arthropods. As we show below, combining both terrestrial and marine arthropod biomasses, we get a total value of **≈1 Gt C**.

Previous estimates

After surveying the literature for an estimate on the global biomass of arthropods, we could not find a reliable source based on sampling of arthropod biomass across biomes. We first give a rough survey of global estimates from the literature. The global number of insects, which are a part of the arthropod phylum, was estimated by Hölldobler & Wilson (194) at $\approx 10^{18}$ individuals based on an earlier estimate by Williams (195) put the total number of insects between 10^{17}-10^{19} based on the density of insects in the soil of southeast England and extrapolated this density to the global terrestrial area of the world. As a result of the International Biological Program, a global analysis of soil fauna, including soil arthropods, was conducted (146). Building on these data, a more recent study (144) added data on several groups of soil arthropods. These two studies, however, do not include all groups of soil arthropods, and not all habitats. In a recent paper estimating the global amount of DNA in the biosphere (196), the authors used a number of 9×10^{18}, presumably based on Williams (195), alongside an estimate they give for the average mass of 10^{-5} kg per organism, to arrive at a global wet biomass of 90 Gt. This is equivalent to ≈13 Gt C when recast as carbon biomass (assuming 30% dry weight and 50% carbon of the dry weight).

Terrestrial arthropods

We took two different approaches to estimate the total biomass of arthropods. The first approach relies on measured biomass densities from the literature which we extrapolate to the global terrestrial ice-free land surface to yield a global biomass estimate. Most measurements of the biomass densities of arthropods we could find in the literature were measured in forests and savannas. Most studies either measure soil arthropod biomass or canopy arthropod biomass. Stork (197) measured the relative contribution of different habitats to the total biomass of arthropods and found litter, soil and canopy arthropods to be the main contributors to the arthropod biomass. Therefore, we combine the estimates for litter (198, 199), canopy (200–202)and soil arthropods (202, 203), each around ≈1 g C m^{-2} and estimate the total biomass density of arthropods at ≈3 g C m^{-2} (link to full calculation). We apply these average biomass densities uniformly across the ice-free terrestrial surface area of $\approx 1.3 \times 10^{14}$ m^2 and thus get an estimate of ≈0.4 Gt C (link to full calculation).

The second approach uses estimates for the average weight of an individual arthropod and estimates for the total number of arthropods in the world to calculate the global biomass of arthropods. Some of the studies which measured the biomass density of arthropods also reported the total individual densities in the same sites (200–202). Dividing the biomass densities by the mass densities gives an estimate of ≈0.1 mg C for the average biomass per individual arthropod (link to full calculation). To get an estimate of the global biomass of arthropods from the average biomass of an individual arthropod, we need information on the total number of arthropods globally. One crude estimate that has been mentioned repeatedly in the literature is that of Williams (195), who estimated 10^{18} insects in the world. Multiplying the average biomass of an individual arthropod with this estimates for the total number of arthropods gives an estimate of 0.1 Gt C (link to full calculation). Our final best estimate for the total arthropod biomass is the geometric mean of biomass estimates we got from the two approaches, which is **≈0.2 Gt C**.

The estimates we get from the two approaches are reassuringly similar. It is possible that the estimates of terrestrial arthropod biomass we provide here are an overestimate of the actual biomass, as our biomass densities are taken mainly from forests, which contain high productivity (and thus also arthropod densities) relative to other biomes. When applying the biomass densities of forests to the entire terrestrial ecosystem, we are likely to overestimate the global biomass of arthropods.

It is noteworthy that for a subset of the groups of terrestrial arthropods, such as termites, a more detailed estimate of biomass exists (204). Using the characteristic biomass densities for this group in each biome as reported in Sanderson (204), and multiplying these densities by the area of each biome (link to full calculation), we get a total of ≈0.07 Gt C, which is in line with the above values by being lower but not a negligible fraction of the 0.25-0.5 Gt C values.
We now analyze the associated uncertainty of the estimate for the total biomass of terrestrial arthropods, which we report as a fold change factor from the mean, representing a range akin to a 95% confidence interval of the estimate. In order to quantify the associated uncertainty, we estimate the biomass of terrestrial arthropods using two independent methods. The first method uses estimates of soil and canopy biomass densities of arthropods mainly in forests and savannas. The uncertainty in those estimates is about ≈2-fold, with a similar within-study uncertainty (link to full calculation). The second method uses an estimate for the average mass of a single arthropod, along with an estimate for the total number of insects. The uncertainty in the average mass of a single arthropod is around 4-fold, while the uncertainty in the estimate of the total amount of arthropods is about an order of magnitude. Combining these two uncertainties together, the uncertainty associated with the second method is ≈15-fold (link to full calculation). As our best projection for the uncertainty associated with the estimate of the biomass to terrestrial arthropods, we take the highest uncertainty of the different uncertainties discussed above. We thus project an uncertainty of ≈15-fold associated with the estimate of the biomass of terrestrial arthropods (link to full calculation). As we stated above, however, our estimate is more likely to be an overestimate than an underestimate.

Marine arthropods

For estimating the total abundance of marine arthropods, we based our estimates on data from the marine ecosystem biomass data (MAREDAT) initiative. This initiative, which quantified global biomass of different plankton groups, is part of the Marine Ecosystem Model Intercomparison Project. The Marine Ecosystem Model Intercomparison Project was initiated in 2007 to facilitate development of dynamic ocean models which include biological information. In the context of this project, it was decided to collect and organize existing biomass concentration measurements for the previously defined “key plankton functional types that need to be simulated explicitly to capture important biogeochemical processes in the ocean” (205). The MAREDAT database contains measurements of the biomass concentration for each plankton group. From this database Buitenhuis et al. (206) make estimates for the global biomass of each plankton group by using a characteristic biomass concentration for each depth (either a median or average of the values in the database) and applying it across the entire volume of ocean at that depth. This approach results in two types of estimates for the global biomass of each plankton group: a so called “minimum” estimate which uses the median concentration of biomass from the database, and a so

called “maximum” estimate which uses the average biomass concentration. Because the distributions of values in the database are usually highly skewed by asymmetrically high values, the median and mean are loosely associated by the MAREDAT authors with a minimum and maximum estimate. The estimate based on the average value is more susceptible to biases in oversampling singular locations such as blooms of plankton species, or of coastal areas in which biomass concentrations are especially high, which might lead to an overestimate. On the other hand, the estimate based on the median biomass concentration might underestimate global biomass as it will reduce the effect of biologically relevant high biomass concentrations. Therefore, here and in all estimates based on MAREDAT data, we take the geometric mean of the “minimum” and “maximum” estimates (actually median and mean values of the distribution) as our best estimate, which will increase our robustness to the effects discussed above. We do not consider the range of the “minimum” and “maximum” estimates as the uncertainty of our estimates, as there are many more sources of uncertainty. We discuss in detail the uncertainties of the estimates based on the MAREDAT database in a dedicated section. The data in the MAREDAT database is divided into plankton size classes: microzooplankton (zooplankton between 5 and 200 µm in diameter), mesozooplankton (zooplankton between 200 and 2000 µm in diameter) and macrozooplankton (zooplankton between 2 and 10 mm). We are interested in the biomass of arthropods in each class. For microzooplankton, we assume the biomass fraction of arthropods is negligible, as the definition of microzooplankton in the MAREDAT dataset excluded copepods, moving copepod biomass to the mesozooplankton category (207). For mesozooplankton, we assume that biomass is dominated by crustaceans such as copepods. We base this assumption on a recent global 18S ribosomal DNA sequencing effort that was part of the *Tara* Oceans campaign (130). First, figure W3 in the companion website of de Vargas et al. (130) shows a good correlation between rDNA copy numbers and biomass of the organism, and gives evidence for the claim that one can reconstruct the relative biomass contribution of several different taxa based on their rDNA read abundance. We use this correlation in conjunction with figure W10 of the companion website of de Vargas et al. (130), which presents the abundance distribution of reads of mesozooplankton between different taxa. From this figure, it is evident that the abundance of mesozooplankton is mostly composed of copepods and protists from the group Rhizaria. Yet Rhizaria are probably not a major fraction of the mesozooplankton in the MAREDAT database as a recent study (208) has indicated that although Rhizaria are a major fraction of mesozooplankton, they are usually under sampled because many of them are delicate and are severely damaged by plankton nets or fixatives used in surveys such as the ones used to build the MAREDAT. Because Rhizaria are usually under sampled in conventional sampling techniques, we assume the samples on which the biomass of mesozooplankton estimate is based, have a small fraction of Rhizaria, and thus are largely dominated by copepods. We therefore use the biomass estimate for mesozooplankton from Buitenhuis et al. (206) which is the basis of the MAREDAT value as an estimate of the copepod biomass. Buitenhuis et al. (206) estimated the biomass of mesozooplankton at around ≈0.4 Gt C (link to full calculation). Data on mesozooplankton biomass in the study was integrated down to a depth of 500 meters. The data reported was standardized to a mesh size of 333 µm, but smaller mesh sizes of 200 µm, which is the lower range of the size fraction of mesozooplankton, will increase the total biomass by ≈1.5-fold (209; link to full calculation). We correct for this effect and thus we use ≈0.6 Gt C as the

global estimate of copepod biomass, assuming that contributions from lower depths will not change the estimate dramatically.
Some arthropods are also included in the macrozooplankton size category (zooplankton between 2 and 10 mm) with species such as Antarctic krill, which Atkinson et al. (210) estimated to be around ≈0.06 Gt C. The "minimum" and "maximum" estimates of the total biomass of macrozooplankton are 0.2 Gt C and 1.5 Gt C, respectively (206). We use the geometric mean of those estimates ≈0.5 Gt C (link to full calculation). Macrozooplankton contains organisms from many phyla such as arthropods, cnidarians, chordates, annelids, molluscs, ctenophores and representatives from Chaetognatha (a phylum of pelagic worms). As described in the section on cnidarians and molluscs, there are databases dedicated specifically to gelatinous plankton (which includes cnidaria, chordata and ctenophora) and to pteropods. The estimate of the combined biomass of pteropods and gelatinous plankton is around ≈0.2 Gt C, which leaves ≈0.3 Gt C of macrozooplankton for the other phyla. Analysis of the data from the MAREDAT database reveals that macrozooplankton biomass is dominated by arthropod biomass. We thus estimate that most of the remaining ≈0.4 Gt C of macrozooplankton is arthropod biomass (link to full calculation). We thus arrive at an estimate of total marine arthropod biomass to **≈1 Gt C** integrating over all relevant size ranges.
In deriving the uncertainty, we note that the estimate of the biomass of marine arthropods is largely based on data present in the MAREDAT database. Some of the sources of uncertainty include possible sampling biases, lack of sampling in many parts of the ocean and uncertainties regarding conversion relations between numbers of individuals to biomass. Due to the presence of many uncertainty sources, most of which are hard to quantify, we took a different approach for projecting the uncertainty of the estimates based on data from the MAREDAT initiative. We further describe our approach in a dedicated section below.

Cnidarians

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

The phylum cnidaria contains both corals and jellyfish. Taken together, these two groups contribute about **≈0.1 Gt C** to the total biomass.
To estimate jellyfish biomass, we use a recent paper by Lucas et al. (211), which has integrated a database using over 90,000 quantitative abundance data for gelatinous zooplankton, together with biometric equations for the conversion of abundance data into carbon biomass. From this data, Lucas et al. calculated a geometric mean of biomass per unit volume and applied it uniformly across the total volume of the upper 200 m of the ocean, which yields an estimate of 0.04 Gt C of gelatinous zooplankton biomass. Of the 38.3 Mt C, 92% are contributed by cnidarians, making their biomass around 0.04 Gt C (link to full calculation). It is important to note that this estimate takes into account only the top 200 m of the water column, and data regarding deeper oceanic zones in scarce. In addition, this estimate takes into account only the pelagic medusa phase of the jellyfish lifecycle, and doesn't include the polyp phase, which is benthic.

Coral biomass was not estimated by Lucas et al. (211) as corals are benthic, and thus we use other approaches to estimate their biomass. We estimate the total biomass of corals in coral reefs by first calculating the total surface area of coral tissue globally, and then convert this value to biomass by the carbon mass density of coral tissue per unit surface area. We estimate the total surface area of corals worldwide using two approaches. The first approach estimates the total surface area of corals using the total area of coral reefs ($\approx 0.25\times10^{12}$ m^2; 212). We estimate that 20% of the reef area is covered by corals (we use ≈20% coverage, a similar value to that of the Great Barrier Reef; 213). This gives us the projected area of corals. Corals have a complex 3D structure that increases their surface area. To take this effect into account, we use a recent study that estimated the ratio between coral tissue surface area and projected area at ≈5 (214). This gives a global coral surface area of $\approx 2.4\times10^{11}$ m^2 (link to full calculation). To convert the total surface area to biomass, we use estimates for the tissue biomass per unit surface area of corals of ≈400 g C m^{-2} (215). This yields a total of ≈0.1 Gt C (link to full calculation). The second approach uses an estimate of the global calcification rate in coral reefs of ≈0.75 Gt $CaCO_3$ yr^{-1} (216). We divide this rate by the surface area specific calcification rate of corals ($\approx 10^4$ g $CaCO_3$ m^{-2} yr^{-1}; 217, 218). This yields a total coral surface area of $\approx 6\times10^{10}$ m^2 (link to full calculation). We multiply the total surface area by the tissue biomass per surface area (≈400 g C m^{-2}; 215) to get ≈0.03 Gt C (link to full calculation). Our final estimate is the geometric mean of both approaches, which is ≈0.05 Gt C (link to full calculation). An important caveat of this analysis is that it doesn't include contribution of corals outside coral reefs, like those located in seamounts. Nevertheless, we account for this biomass of corals which are out of formal coral reefs when calculating the total benthic biomass.

Molluscs

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Molluscs are a large phylum of marine invertebrates, and contains several classes, such as gastropods (snails, slugs, and pelagic sea hares and sea butterflies), cephalopods (squids and octopuses) and bivalves (clams, oysters etc.). Out of all the different species, only gastropods are both marine and terrestrial, and mostly gastropods and cephalopods are pelagic.
As part of the marine ecosystems biomass data initiative, quantifying the global biomass of different plankton groups, Buitenhuis et al. (206) constructed a database containing samples of biomass of pteropods, which are a group of pelagic gastropods including sea butterflies and sea angels. From this database Buitenhuis et al. (206) generated estimates for the global biomass of pteropods by using a characteristic biomass concentration for each depth (either a median or average of the values in the database) and applying across the entire volume of ocean at that depth. The estimates based on median and average biomass concentrations at each depth are 0.026 Gt C and 0.67 Gt C, respectively (206). We use the geometric mean of the two estimates to arrive at ≈0.15 Gt C (link to full calculation). The data contains measurements down to 2000 m depth, and thus doesn't include the entire depth range of the ocean. Nevertheless, the depth profiles of the data show that abundance drops with depth so that biomass of lower layers is smaller than surface water biomass. Our best estimate of cephalopod biomass is ≈0.05 Gt C,

based on Rodhouse & Nigmatullin (219). We thus took a value of **0.2 Gt C** as the estimate of the global biomass of molluscs under the assumption that the biomass contribution of other types of molluscs such as benthic molluscs or terrestrial gastropods is minor compared to the contribution of pteropods.

Protists

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Protists are a broad and highly diverse group of eukaryotes, present in the marine, terrestrial and deep subsurface environments.

Marine protists

The estimate of marine protists is based on the MAREDAT initiative, where we included estimates of all plankton groups that are dominated by protists. The main groups with a significant biomass contribution were picoeukaryotes (206, 220), microzooplankton (206; defined not to include copepod biomass), diatoms (221), and *Phaeocystis (206)*. The MAREDAT database contains measurements of the biomass concentration for each of plankton groups above. Members of the MAREDAT initiative used this database to estimate the global biomass of each plankton group by using a characteristic biomass concentration for each depth (either a median or average of the values in the database) and applying across the entire volume of ocean at that depth. Two types of estimates are supplied for the global biomass of each plankton group: a "minimum" estimate which uses the median concentration of biomass from the database, and a "maximum" estimate which uses the average biomass concentration. Because the distributions of values in the database are usually highly skewed by asymmetrically high values the median and mean are loosely associated by the authors of the MAREDAT study with a minimum and maximum estimate. The estimate based on the average value is more susceptible to biases in oversampling singular locations such as blooms of plankton species, or of coastal areas in which biomass concentrations are especially high, which might lead to an overestimate. On the other hand, the estimate based on the median biomass concentration might underestimate global biomass as it will reduce the effect of biologically relevant high biomass concentrations. Therefore, here and in all estimates based on MAREDAT data, we take the geometric mean of the "minimum" and "maximum" estimates (actually median and mean values of the distribution) as our best estimate, which will increase our robustness to the effects discussed above. We do not consider the range of the "minimum" and "maximum" estimates as the uncertainty of our estimates, as there are many more sources of uncertainty. We discuss in detail the uncertainties of the estimates based on the MAREDAT database in a dedicated section.

For picoeukaryotes, Buitenhuis et al. (206, 220) estimates that they represent ≈60% (49-69%) of the global biomass of picophytoplankton. Estimates for the global picophytoplankton biomass (206, 220) converge at ≈0.4 Gt C which translates to ≈0.2 Gt C of picoeukaryotes (link to full calculation). Picoeukaryotes contain both protists and plant species (like chlorophytes). It seems that, from the available literature, the biomass distribution between them is not strongly favored

towards one class (222), we thus chose to estimate the protist fraction at ≈0.1 Gt C (link to full calculation).

Protists in the picoplankton to nanoplankton size range (0.8-5 µm in diameter) include not only autotrophic, but also heterotrophic organisms. As we could not find a reliable resource for estimating the biomass of heterotrophic pico-nanoplankton we use a recent global 18S ribosomal DNA sequencing effort that was part of the *Tara* Oceans campaign (130). Based on this study, it appears that in the locations sampled the biomass of heterotrophic protists in this size range is about twice that of autotrophic protists, so we estimate a biomass of ≈0.3 Gt C of heterotrophic pico-nanoplankton protists. Relying on 18S sequence abundance as a proxy for biomass is not a well-established practice, and has various biases, but for lack of any other alternative we could find to perform the estimate, we chose to use it. Yet, we note that this plays a minor role in our analysis that in any case will not affect any of the major conclusions of our study.

For microzooplankton, estimates in Buitenhuis et al. (206) are of ≈0.6 Gt C (link to full calculation). For diatoms, Leblanc et al. (221) estimates a global biomass of ≈0.3 Gt C (link to full calculation). For *Phaeocystis*, the estimate in Buitenhuis et al. (206) is ≈0.3 Gt C, but the data has been noted to lack coverage and to have a bias to coastal environments (link to full calculation). As stated in Buitenhuis et al. (206), the data from the MAREDAT initiative doesn't contain the biomass of nanophytoplankton (phytoplankton between 2 and 20 µm) and autotrophic dinoflagellates. Nevertheless, this omission might be compensated by overestimation of *Phaeocystis* biomass because of sampling bias, so overall the sum of all the different phytoplankton fits well with total chlorophyll measurements from the WOA 2005. This is why we chose to take the total value of 1.5 Gt C as the sum of protist biomass from the MAREDAT initiative.

For phytoplankton, biomass and number of observations drop rapidly below 200 meters, and a recent paper (223) estimated that the integrated cell abundance in the deep ocean (2000-4000 m) is more than an order of magnitude smaller than for the top 200 meters. Therefore, our estimates probably will not be affected significantly by omission of deep sea plankton biomass. Also for heterotrophic protists, cell abundance in the deep ocean is much lower, with estimates for their total biomass of ≈0.1 Gt C (132). Other groups of protists, such as coccolithophores and foraminifera, for which data exist in the MAREDAT initiative datasets, were not included because their relative biomass contribution is an order of magnitude smaller.

In addition, a recent paper by Biard et al. (208) measured the contribution of protists from the Rhizaria super-group. As stated in the marine arthropod section, the biomass of this group of protists is underrepresented in conventional plankton estimates, such as that of the MAREDAT initiative because of the delicate nature of the organisms in the group. Biard et al. use in-situ imaging to estimate the biomass concentration of Rhizaria. Biard et al. divided the data into three depth layers (0-100 m, 100-200 m, and 200-500 m), and multiplied median biomass concentrations at each depth layer across the entire volume of water at that layer to generate a global estimate. The resulting estimate for the biomass of Rhizaria in the top 500 meters of the

ocean is ≈0.2 Gt C (link to full calculation). As with other estimates, because of lack of sampling below 500 this might be an underestimate of actual Rhizaria biomass.

An additional location in which protists reside is attached to particulate organic matter. To estimate the biomass of particle-attached protists in the marine environment, we rely on a several studies that have measured the relative number of particle-attached protists and prokaryotes at the epipelagic (41, 224), mesopelagic (41) and bathypelagic (50) layers. For each study, we calculate the characteristic ratio between the number of protist and prokaryote cells. To estimate the biomass of particle-attached protists relative to prokaryotes, we use estimates on the carbon content of protists in the epipelagic, mesopelagic and bathypelagic layers (132, 224). We compare this carbon content to our best estimate of the carbon content of particle-attached prokaryotes (see marine bacteria and archaea section). We generate estimates for the biomass of particle-attached protists relative to particle-attached prokaryotes in the epipelagic, mesopelagic and bathypelagic layers. Overall, we estimate particle-attached protists have about the same biomass as particle-attached prokaryotes (link to full calculation). We estimate ≈0.3 Gt C of particle-attached prokaryotes in marine snow, thus leading to an estimate of ≈0.3 Gt C of particle-attached protists. Estimates of photosynthetic benthic protists, the kelps growing on rocky substrates put their the global standing crop of 0.02 Gt C (225). A similar value can be calculated from annual productivity given by De vooys (16) assuming one turnover of the standing crop each year (18).

The estimate of the biomass of marine protists is largely based on data present in the MAREDAT database. There are many sources of uncertainty associated with the data collected in the database, which include possible sampling biases, lack of sampling in certain parts of the ocean and uncertainties regarding conversion relations between numbers of individuals to biomass. Due to the presence of many uncertainty sources, some of which are hard to estimate, we took a different approach for projecting the uncertainty of the estimates based on data from the MAREDAT initiative. For a detailed characterization on the uncertainty of the data in the database, see the relevant section below. Overall, bringing together estimates from both the MAREDAT initiative data and biomass estimates of Rhizaria and particle-attached protists, we get a total of **≈2 Gt C** for marine protists.

Terrestrial protists

After searching the literature, we could not find a comprehensive account of the biomass of protists in soils. We thus generated a crude estimate of the total biomass of protists in soil based on estimating the total number of individual protists and the characteristic carbon content of individual protists. Generally, protists are divided into four morphological groups, i.e. flagellates, ciliates, naked amoebae and testate amoebae (226). The relative proportions of these morphotypes vary greatly between studies. Not all studies report values for each of the groups. To estimate the total number of soil protists, we collected data from 42 independent studies which measured the biomass density of protists per gram dry weight of soil (227–248). We specifically avoided using studies which rely on culturing of soil samples and the most probable number method, as these methods are not reliable (248).

To go from the measurements reported in these studies to an estimate of the total number of soil protists, we bin the measurements based on the habitat in which they were measured. For studies which reported more than one value in a specific habitat, we first calculate the mean of these values. We then calculate the mean value from the different values reported in different studies within the same habitat. We thus generate characteristic values for the number of individual protists per gram soil in each habitat and for each group of protists. Missing values in specific habitats were filled based on similar habitats (e.g. for boreal forests we use values from temperate forests) or based on general values from expert assessments of the characteristic values for each group of protists (243). We used two approaches to calculate the mean values for measurements within the same study and between studies. We either used the arithmetic mean or the geometric mean of values as explained below. The estimate based on the arithmetic mean is more susceptible to sampling bias, as even a single measurement which is not characteristic of the global population (such as samples which are contaminated with organic carbon sources, or samples which have some technical biases associated with them) might shift the average number of individuals significantly. On the other hand, the estimate based on the geometric mean might underestimate global biomass as it will reduce the effect of biologically relevant high biomass concentrations. As a compromise between these two caveats, we chose to use as our best estimate the geometric mean of the estimates using the two methodologies.

To convert measurements of the total number of protists per gram soil to units of number of individuals per m^2, we need two parameters - the bulk density of soil, and the depth at which the measurements were conducted. For the bulk density of soil, we use a global mean of ≈ 1.2 g cm^3 (249). The mean depth in which samples were collected (across studies) is ≈8 cm. We use this value as the characteristic value for the depth at which the measurements were conducted (link to full calculation). By multiplying the total number of individual protists per gram of soil by the bulk density of soil and the depth of sampling, we get an estimate of the total number of protists of each morphological group in each habitat in the top 8 cm of soil. These estimates do not take into account biomass contributions from soil layers below 8 cm. To account for these additional contributions, we estimate the fraction of the total biomass of soil protists found in the top 8 cm of the soil profile out of the full depth of soil. To estimates the fraction of the total biomass of soil protists found in the top 8 cm of soil profile, we rely on two studies (110, 144) which provide estimates for the distribution of microbial biomass along the soil profile. We use the geometric mean of those two studies, and estimate that ≈50% of the total biomass of soil protists is found in the top 8 cm of the soil profile (link to full calculation). Multiplying our estimates for the total number of soil protists by a factor of two gives us an estimate for the total number of soil protists from each morphological type per m^2 in each habitat. To calculate the global number of protists from each morphological type, we multiply the estimates for the total number of individuals per m^2 in each habitat by the total area of each habitat (23). We arrive at an estimate of $\approx 1.5\times10^{22}$ ciliates, $\approx 3\times10^{25}$ flagellates, $\approx 1.5\times10^{24}$ naked amoebae and $\approx 4\times10^{23}$ testate amoebae globally (link to full calculation).

In addition to estimating the total number of each morphological type of soil protist, we also estimate the characteristic carbon content of each morphological type. We collected data from 11 studies which measured the carbon content of the different morphological types (238, 239, 246,

250–257). For studies which report more than one measurement, we first calculate the geometric mean of values within the study. We then calculate the geometric mean of values from different studies for each morphological type. We arrive at an estimate of ≈900 pg C for a ciliate cell, ≈20 pg C for a flagellate cell, ≈200 pg C for a naked amoebae cell, and ≈1500 pg C for testate amoebae (link to full calculation). We multiply the total number of cells of each morphological type by the characteristic carbon content of each morphological type, and sum over all morphological types to arrive at an estimate of the total biomass of soil protists, which amounts to **≈1.5 Gt C** (link to full calculation).

We now analyze the associated uncertainty of the estimate for the total biomass of terrestrial protists, which we report as a fold change factor from the mean, representing a range akin to a 95% confidence interval of the estimate. We first assess the uncertainty associated with the estimate of the total number of individual protists. We derive the intra-study uncertainty for the values reported in studies which reported more than one measurement (link to full calculation). We also calculate the inter-study uncertainty for different studies within the same habitat. We use the maximum of the intra-study and inter-study uncertainty as our projection of the uncertainty associated with the estimate of the total number of protists in each habitat (link to full calculation). We integrate the uncertainties associated with our estimates of the bulk soil density and the fraction of the total biomass of soil protists found in the top 8 cm of the soil profile. We could not find an estimate for the uncertainty around the global average of the the bulk soil density, so we crudely use ≈2-fold as our best projection. We use the 95% confidence interval around our mean fraction of the total biomass of soil protists found in the top 8 cm as our projection for the uncertainty associated with this parameter, which is ≈3-fold (link to full calculation). We propagate the uncertainties for the total number of protists in each biome to the estimate of the total number of protists for each protists group. This yields a projection of ≈6-fold uncertainty for the total number of ciliates, ≈5-fold uncertainty for the total number of flagellates and testate amoebae, and ≈4-fold for the total number of naked amoebae (link to full calculation). For the carbon content of each morphological type, we assess the intra-study uncertainty and interstudy uncertainty, and use the maximum of both as our projection of the uncertainty associated with the characteristic carbon content of each morphological type. We project an uncertainty of ≈3-fold for the characteristic carbon content of ciliates and flagellates, ≈8-fold uncertainty associated with the estimate of the carbon content of naked amoebae and ≈2-fold uncertainty associated with the estimate of the carbon content of testate amoebae (link to full calculation). Finally, we combine the uncertainties for the total number of cells and the characteristic carbon content of each morphological type and propagate the uncertainty to the total estimate of the biomass of soil protists. We thus arrive at a projection of ≈4-fold uncertainty associated with our estimate of the total biomass of soil protists (link to full calculation).

Deep subsurface

Another potential contribution to the global biomass of protists comes from terrestrial and marine deep subsurface environments. The biomass of protists in both environments has not been extensively studied. For marine deep subsurface, eukaryotes are probably dominated by fungi (258). As we estimate the biomass of fungi to be very small in those environments (see fungi section), we estimate that the biomass of other eukaryotic microbes such as protists will also be

low. For the terrestrial deep subsurface, most early estimates of the abundance of protists in aquifers were based on cultivation-based methods, which are likely not to accurately represent the distribution of protists (136, 259). Based on these measurements, researchers have claimed the distribution of protists is limited to shallow aquifers (260), with the abundance of protists in deeper aquifers being orders of magnitude lower than that of prokaryotes (136). Nevertheless, more recent findings have identified the presence of protists in deep terrestrial subsurface (261). Borgonie et al. have claimed the abundance of protists in their samples is low but provided no quantitative account of the abundance of protists which allows estimation of their global abundance. Measurement of the abundance of protists in uncontaminated aquifers using direct microscopic counts are very scarce and are focused on shallow aquifers (262, 263), which do not reflect the abundance of deep subsurface aquifers. Protists might also be present in the oceanic crust, but at present we could not find reliable sampling of their abundance in this environment. Due to the lack of data on the abundance of protists in the terrestrial deep subsurface and oceanic crust, we do not include them in the current analysis. We highlight this knowledge gap in the discussion section of the paper main text.

Sanity checks on the MAREDAT dataset

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

A major source of information regarding the global biomasses of marine taxa is the MAREDAT database. As stated in the specific sections relying on data from the database, there are many sources of uncertainty associated with the estimates stemming from the MAREDAT data. As some of these uncertainties are hard to quantify accurately, we chose a different route to projecting the total uncertainty of the estimates in the MAREDAT database. We compare the estimates generated based on the MAREDAT database with estimates based on independent sources as a consistency check for the robustness of the biomass estimates we use in this study, and to quantify the uncertainty of our estimate.

Comparison of MAREDAT and the *Tara* oceans dataset

The *Tara* Oceans expedition is a campaign set out by the multinational Tara Oceans consortium, which sampled microscopic plankton at 210 sites and at depths up to 2000 m in all the major oceanic regions during expeditions from 2009 through 2013 (264). One of the products of this expedition was a study which described the diversity of eukaryotic plankton across the global oceans (130). The dataset this study generated divides the plankton community based on size ranges (pico-nano-, nano-, micro- and meso-plankton). The data in the Tara Oceans provides only number of reads for each taxon. The fraction of reads that a taxon has out of the total number of reads can be used as a proxy for the biomass fraction of the taxon, but not as a proxy of its absolute biomass. Relying on 18S sequence abundance as a proxy for biomass is not a well-established practice, and has its own biases, but we chose to use it for the sake of comparing it to independent approaches such as the MAREDAT database. Each plankton size fraction sampled in the study was sequenced to approximately the same sequencing depth (≈120 million reads). This means that the 18S read data can provide a possible proxy for the biomass fraction of a certain

taxon within a size fraction, but not across size fractions. The *Tara* Oceans dataset was collected independently of the biomass estimates of the MAREDAT database and can thus serve for cross validation. Our goal is to compare those two datasets and see if they agree with each other. Therefore, to compare those two datasets we need to convert the data in the MAREDAT database to biomass fractions. We focus this comparison on two test cases: the biomass of diatoms and the total biomass of nanoplankton and microplankton.

We begin by describing how to compare the biomass estimates of diatoms based on the Tara Oceans dataset and the MAREDAT dataset. As the data in those two datasets is structured differently, we first need to make corrections to the data so a valid comparison will be available. In the Tara Ocean dataset, diatoms appear mainly in the nanoplankton (5-20 µm in diameter) and microplankton (20-180 µm) size fractions. In order to make a comparison to the MAREDAT database we need to find the corresponding groups in the MAREDAT database. The corresponding groups in the MAREDAT database are the microzooplankton and the diatom groups (zooplankton between 5 and 200 µm in diameter)[1]. The total biomass of the diatom and microzooplankton groups in the MAREDAT database is ≈1 Gt C with diatoms accounting for about 30% of this biomass (0.3 Gt C). The microzooplankton biomass estimates in the MAREDAT database do not include copepods, which were moved to the mesozooplankton group. Fragile protists such as Rhizaria, are probably also undersampled in the MAREDAT database. Therefore, to correct for these effects such that we could compare the MAREDAT and *Tara* Oceans datasets, we remove metazoa (dominated by arthropods) and Rhizaria reads from the relevant size fractions in the *Tara* Oceans dataset (nano and microplakton). After correcting for those biases, the biomass fraction of diatoms in microplankton in the Tara Oceans dataset is between 16%-33%, which fits well the estimate of ≈30% based on the MAREDAT data (link to full calculation).

We also compared the MAREDAT and *Tara* Oceans datasets, by deriving an independent estimate of the total biomass of nanoplankton and microplankton. To generate this estimate, we begin with the independently measured biomass of Rhizaria. The independent measurement using microscopy by Biard et al. (208) has estimated ≈0.2 Gt C of Rhizaria above 600 µm in diameter. We assume that this biomass represents the biomass of Rhizaria in mesozooplankton. As we derive in the marine arthropod section, Rhizaria represent ≈40% of the total mesoplankton biomass (link to full calculation). The remaining 60% are made up mainly of arthropods. This would put the total mesozooplankton arthropods biomass at ≈0.3 Gt C. The total arthropod biomass based on the MAREDAT database is estimated to be ≈0.5 Gt C in the sum of the nano, micro and mesozooplankton size fractions, which leaves ≈0.2 Gt C of nano and microzooplankton arthropod biomass. Based on the *Tara* Oceans data, the nano and microzooplankton arthropod biomass accounts for ≈50% of the total nano and microplankton biomass. We use the estimate we just calculated of ≈0.2 Gt C of arthropods in the nano and

[1] The biomass of *Phaeocystis* contains also cells in the relevant size range. Nevertheless, we exclude its contribution here due to two main reasons. First, the *Phaeocystis* group contains also cells above 200µm diameter. In addition, several lines of evidence imply that the biomass of *Phaeocystis* is probably overestimated in the MAREDAT dataset (based on MAREDAT paper and the small fraction of archaeplastida in the *Tara* database). Therefore, we exclude this biomass from this specific analysis.

microplankton size fractions and combine it with the estimate of the biomass fraction of arthropods in the nano and microplankton size fractions from the Tara Oceans dataset. This yields an estimate for the total nano and microplankton biomass which is about ≈0.5 Gt C (link to full calculation). The biomass of nano and microplankton is estimated at ≈1 Gt C based on the MAREDAT database, which is about 2-fold larger than the estimate we got based on combination of information from Biard et al. the Tara Oceans dataset and the MAREDAT database.

Cyanobacteria

For cyanobacteria, a study based on independent data of cell concentrations (265) estimated the total number of cells of prochlorococcus and synechococcus, which are estimated at $\approx 3\times10^{27}$ and $\approx 7\times10^{26}$ cells, respectively. We assume those two cyanobacterial clades represent a dominant part of the global biomass of cyanobacteria. We use the estimate of the total number of those two clades, in conjunction with estimates for the carbon content of a single prochlorococcus and synechococcus cell (220), which are ≈40 fg cell^{-1} and ≈250 fg cell^{-1}, respectively. Multiplying the number of cells by the average carbon content of a cell, we estimate the total biomass of cyanobacteria at about ≈0.3 Gt C. We compare this estimate to the estimate of the biomass of cyanobacteria based on the MAREDAT database. Buitenhuis et al. (220) estimate, based on the MAREDAT database, the total biomass of picophytoplankton at around ≈0.3 Gt C (link to full calculation). Out of this biomass, Buitenhuis et al. estimate ≈31-51% are contributed by cyanobacteria. This suggests that the estimate for the total biomass of cyanobacteria based on the MAREDAT database is ≈0.2 Gt C. Thus, these two independent estimates are less than 2-fold apart.

Comparison of phytoplankton to remote sensing measurements

The biomass of all phytoplankton can also be assessed based on remote sensing data, which calculates the global depth-integrated chlorophyll biomass. Reports by Antonine et al. (266) and Behrenfeld & Falkowski (267), which are based on remote-sensing, put global phytoplankton biomass at a similar value of ≈0.75 Gt C. Summing up all the biomass contributions of photosynthetic organisms in the MAREDAT database, we get ≈1.3 Gt C, which is about 2-fold higher but within the same order of magnitude as estimates from remote sensing (link to full calculation).

Most of the differences between the estimates based on the MAREDAT database and the independently calculated estimates used for the consistency checks are below an order of magnitude. Nevertheless, the independent estimates we derived are probably not entirely independent of the estimates based on the MAREDAT database, as some of them rely partially on data which is used also in MAREDAT-based estimates. Taking into consideration all the different consistency checks and caveats associated with the MAREDAT database, we crudely estimate the uncertainty of the estimates based on the MAREDAT database, which we report as a fold change factor from the mean representing a range akin to a 95% confidence interval of the estimate, to be about one order of magnitude.

Comparison of the biomass of producers and consumers

In Fig. *2C*, we analyze the distribution of biomass between producers (autotrophs) and consumers (heterotrophs) in the terrestrial and marine environments. In the terrestrial environment, we use the biomass of plants as representing the total biomass of terrestrial autotrophs. The remaining terrestrial biomass, including soil bacteria (≈7 Gt C), soil archaea (≈0.5 Gt C), soil fungi (≈12 Gt C), soil protists (≈1.5 Gt C), and terrestrial animals (≈0.5 Gt C) are considered as terrestrial heterotrophs. In the marine environment, the biomass of seagrasses (≈0.1 Gt C), macroalgaea (≈0.1 Gt C), picoplankton (≈0.4 Gt C), diatoms (≈0.3 Gt C) and *Phaeocystis* (≈0.3 Gt C) were summed to produce an estimate of the total biomass of marine autotrophs of ≈1.3 Gt C. This estimate is about 2-fold higher than estimates of the total autotrophic biomass made by Antonine et al. (266) and Behrenfeld & Falkowski (267), based on remote sensing. The remaining marine biomass of ≈5 Gt C including marine bacteria (≈1.3 Gt C), archaea (≈0.3 Gt C), fungi (≈0.3 Gt C), heterotrophic protists (≈1.1 Gt C) and animals (≈2 Gt C), was considered as marine heterotrophic biomass. We consider all heterotrophic organisms as consumers, and thus we include detritivores as part of the consumer biomass. We note that even if we will consider only animal biomass as consumers, the marine animal biomass (≈2 Gt C) is higher than the biomass of all marine producers.

Viruses

We provide a fully detailed calculation for the following analysis, including all of the data as well as the steps taken to achieve these results, in the following link.

Viruses are often introduced as the most abundant biological entity in terms of the number of individuals in the world. To estimate their global biomass, we first focus our efforts on bacteriophages. An often-quoted rule of thumb is that bacteriophages are ≈10 times more numerous than their bacterial and archaeal hosts. Recent studies have revisited and broadly given further support for this claim in the marine environment (268), in the marine deep subsurface (269) and in the terrestrial deep subsurface (270). We can thus use the specific relation between the number of viruses and prokaryotes described in each environment, along with our estimates for the number of prokaryotes in each environment (see relevant sections in the *SI Appendix*) to estimate the total number of bacteriophages. Alternatively, we can also extrapolate the total number of bacteriophages from direct measurements of the concentration of bacteriophages particles. We use the later approach in cases where no reliable data regarding the ratio between bacteriophages and viruses are available, such as in soils, or in cases in which there is enough direct data on bacteriophage concentrations for a robust estimate that is not dependent on the estimate for the total number of prokaryotes.

We proceed to describe the procedure for estimating the total number of bacteriophages in each environment. For the marine environment, we relied on data in Wigington et al. (268) which provided ≈6000 samples of virus concentration in the ocean. We binned the data along the depth of the ocean. For each bin we calculated the mean concentration of viruses. We then multiplied the mean concentration of viruses in each bin by the total volume of that bin to calculate the total

number of viruses in each bin. Finally, we summed up the estimates for the total number of viruses in each bin to calculate the total number of viruses in the ocean. Our best estimate for the total number of viruses in the ocean is $\approx 2\times10^{30}$ (link to full calculation). For the marine deep subsurface environment, we use two recent studies (269, 271) which counted the number of prokaryotes and virus particles in five different sediment types, which cover a broad range of characteristics such as sedimentation rates, total organic carbon, sediment age, and microbial activity. To estimate the total number of bacteriophages in the marine deep subsurface, we used the geometric mean of ratios between the number of virus particles and the number of bacteria and archaea. The geometric mean of the ratios is ≈ 12 (link to full calculation). As noted in the marine deep subsurface prokaryotes section, our best estimate for the number of bacterial and archaeal cells in the marine deep subsurface is $\approx 4\times10^{29}$ cells. Using the ratio between viruses and prokaryotes in Engelhardt et al. (269), we estimate a total of $\approx 5\times10^{30}$ viruses in the marine deep subsurface (link to full calculation). For the terrestrial deep subsurface, the relation between number of virions and cells was measured in only a handful of locations. In addition, measurements of viral concentrations were done in aquifer waters, and thus it is not clear what is the concentration of phages attached to surfaces. The number of attached cells in the terrestrial deep subsurface is claimed to be much higher than the number of suspended cells in aquifers, and it is not clear if this claim holds true also for viruses. Therefore, we generated several estimates of the total amount of viruses in the terrestrial deep subsurface. We used five types of estimates for the virus to prokaryotes ratio. The first estimate is based on assuming a 10:1 ratio. As an alternative estimate we used the relation between the number of viruses and prokaryotes measured in the marine deep subsurface (269). We used this relation along with the concentration of prokaryotes at each depth in the terrestrial subsurface, taken from McMahon & Parnell (98). As a third estimate for the ratio between viruses and prokaryotes in the terrestrial deep subsurface we used the relation between the number of viruses and prokaryotes measured in the terrestrial deep subsurface by Kyle et al. (270). Our fourth estimate is based on an average ratio of 3 between viruses and prokaryotes measured by Pan et al. (272). The fifth estimate is based on an average ratio of 2 between viruses and prokaryotes measured by Roudnew et al. (273). In order to use these five estimates for the ratio between viruses and prokaryotes, we need to feed them with the number of prokaryotes in the terrestrial deep subsurface. We plug into these five estimates our estimate for the number of prokaryotes in groundwater (see the section on the biomass of terrestrial deep subsurface prokaryotes for details on the calculation of the total number of prokaryotes in groundwater). We thus generate five estimates for the total number of bacteriophages in groundwater. It is not clear whether these relations between the concentration of bacteriophages and prokaryotes refer only to unattached cells or includes also attached cells. The total number of bacteriophages we estimated using the five methodologies detailed above take into account only unattached bacteriophages (in groundwater). In order to take into account the possible contribution in groundwater of bacteriophages which are not planktonic, we use a scaling factor that will convert our estimate for the total number of bacteriophages in groundwater to an estimate for the total number of bacteriophages in the terrestrial deep subsurface. As our best estimate for this scaling factor, we use a geometric mean of three estimates. The first takes into account only bacteriophages in groundwater (and thus the scaling factor is 1), and the other two assume a ratio between attached to unattached bacteriophages which are similar to the ratios between attached and unattached cells in the terrestrial deep subsurface. The ratios used for the

other two estimates are 10^2 and 10^3 which are the lower and higher bounds for the ratios between the number of attached and unattached cells reported in McMahon & Parnell (98). We multiply our five estimates for the number of bacteriophages in groundwater by the scaling factor to generate five estimates for the total number of bacteriophages in the terrestrial deep subsurface. We take the geometric mean of the different estimates, which is $\approx 2\times10^{30}$, as our best estimate for the total number of viruses in the terrestrial deep subsurface (link to full calculation). As for soil viruses, after searching the literature, we could find only a handful of sources for the abundance of viruses in soils. The values found are about $\sim10^8$-10^9 virions per gram of soil (274, 275), which translates to $\sim10^{14}$-10^{15} viruses per m^{-3}. Applying this value across the entire terrestrial surface and assuming a mean soil depth of 10 meters (276), we get to an estimate of $\sim10^{29}$-10^{30} virions in soil globally. This is probably an overestimate of the total abundance of bacteriophages in the soil as they are probably not as abundant in lower layers as in the uppermost layers of soil, because of the lower cell abundances in those layers. We use the geometric mean of the lower and upper range of our estimate, which is $\approx 6\times10^{29}$ as our best estimate for the number of viruses in the soil (link to full calculation).

Viruses are also present in fluids inside the oceanic crust (277). This habitat is much less explored than subseafloor sediments, and thus estimates of the total abundance of viruses in the ocean's crust are more speculative. Nigro et al. (277) estimate that viruses are about an order of magnitude more abundant than bacteria and archaea also in the oceanic crust. We estimate a total of 2×10^{28}-2×10^{29} bacterial and archaeal cells in the oceanic crust (see marine deep subsurface prokaryotes section). Thus, we can crudely estimate 2×10^{29}-2×10^{30} viruses in the oceanic crust. Including the contribution from this environment will thus not change our results dramatically, as even the higher estimate is only ≈20% of our estimate for the global number of viruses. Due to the extreme scarcity of data on the abundance of viruses in the oceanic crust, we do not include them in the current analysis. We highlight this knowledge gap in the discussion section of the paper main text.

We combine the number of viruses in each environment and generate an estimate for the total number of viruses of $\approx10^{31}$ (link to full calculation). Because of the vast dominance of bacteria over eukaryotes in terms of number of cells, it is probable that bacteriophages dominate the abundance of all viruses. Bacteriophages have a characteristic capsid diameter of ≈50 nm (278; link to full calculation). Though some of the largest viruses known are ~1000 times larger in volume than a typical bacteriophage, we expect them to be less abundant by a much larger factor and thus will not significantly change the dominance of bacteriophages over eukaryotic viruses biomass. To estimate the total biomass of viruses, we multiply the total number of viruses by a characteristic mass of a virus (in our case, a bacteriophage). To estimate the characteristic mass of a bacteriophage, we rely on a biophysical model of the elemental composition of a virion (279). Plugging into the formulas detailed in (279) a characteristic radius of 25 nm (based on a characteristic diameter of 50 nm; 278), we estimate that a characteristic bacteriophage contains about ≈0.02 fg C (link to full calculation). To estimate the total biomass of viruses, we multiply the characteristic carbon content of ≈0.02 fg C by our estimate for the total number of viruses of $\approx10^{31}$ and arrive at a total biomass of **≈0.2 Gt C** (link to full calculation).

As a consistency check for the estimate of viral biomass, we inspect the ratio between the global viral and bacterial biomasses. Our estimated biomass for viruses is approximately ≈0.1% of the total bacterial biomass. We use the fact that viruses are dominated by bacteriophages, which have a characteristic diameter of ≈50 nm. The rule of thumb diameter of a bacterium is on the order of 1 μm. This means that in terms of volume, viruses will have ≈$1:10^4$ of the volume of a bacterium. Even if highly packed, the dry mass density of viruses cannot increase by more than 2-fold above bacteria which are already ⅓ dry matter by mass. Viruses are on average about 10 times more abundant than bacteria (268), which means that the global biomass of viruses is indeed 1% or less of bacterial biomass, even when considering the fact that most deep subsurface bacteria are up to an order of magnitude smaller in mass than the standard lab bacterium.

We now analyze the associated uncertainty of the estimate for the total biomass of viruses, which we report as a fold change factor from the mean representing a range akin to a 95% confidence interval of the estimate. The main parameters affecting the estimate are the number of bacterial cells in each environment, the ratio between bacterial cells and virions, the size of the virions and their carbon content. Each environment in which bacteria are abundant (namely, marine, soil, marine deep subsurface and terrestrial deep subsurface) has its own ranges of uncertainty regarding the number of cells and the ratio between viruses and bacteria. We survey the different ranges of uncertainty of the number of bacterial cells and the ratio between viruses and bacteria. Then, we discuss the uncertainties in the sizes of viruses and their carbon content.

For soils, there are no good indicators for the uncertainty of our estimate of the total number of phages. The range of ~10^8–10^9 phages per gram of soil introduces an uncertainty of about an order of magnitude. The specific values of the bulk density of soils, as well as the depth of soils also have uncertainty associated with them, which is hard to quantify. Our estimate is likely to be an overestimate, as it likely that the concentration of phages in deeper soil layers will be lower than in shallower layers, as is for prokaryotes. We thus project very crudely an uncertainty of one and a half orders of magnitude associated with our estimate of the number of phages in soils. Because the total number of viruses in soils is relatively small (our best estimate for the total number of viruses in soils is ≈6×10^{29}, which is ≈6% of the total number of viruses we estimate worldwide), we expect the uncertainty associated with the total number of viruses in soils not to affect our estimate of the total number of viruses significantly.

For the terrestrial deep subsurface, our estimate relies on three main factors. The first factor is the ratio between the concentration of viruses and prokaryotes. The second factor is the total number of prokaryotes we plug into the ratio between viruses and prokaryotes in order to estimate the total number of bacteriophages. The third factor is the scaling factor for the total number of viruses in groundwater and the total number of the number of viruses in the terrestrial deep subsurface. For the uncertainty associated with the ratio between the concentration of viruses and prokaryotes, we relied on five different ratios from different sources, as detailed above. We rely on the variability between these five different ratios to calculate the uncertainty associated with the ratio of the number of viruses and bacteria in the terrestrial deep subsurface. We arrive at an uncertainty of ≈2-fold associated with the ratio of viruses and prokaryotes in the terrestrial deep subsurface (link to full calculation). For uncertainty associated with the total number of

prokaryotes we plug into the relation between the concentrations of viruses and prokaryotes, we propagate our uncertainty about the number of prokaryotes to the estimate of the total number of bacteriophages. We use two estimates for the total number of prokaryotes in groundwater - one based on arithmetic mean concentration of cells at each depth bin and one based on geometric mean concentration of cells. We calculate the amount of uncertainty using each of those estimates introduces into the final estimate of the total number of bacteriophages. We arrive at an uncertainty of ≈2-fold (link to full calculation). For the scaling factor between the number of viruses in groundwater and the total number of viruses in the terrestrial deep subsurface, we calculate the 95% confidence interval of the geometric mean scaling factor as a measure of the uncertainty associated with the scaling factor. We arrive at an uncertainty of ≈50-fold associated with the scaling factor (link to full calculation). Combining the uncertainty associated with the three different factors, we calculate a total uncertainty of ≈60-fold associated with our estimate of the total number of viruses in the terrestrial deep subsurface (link to full calculation).

For the number of bacteriophages in the marine environment and in the marine deep subsurface, could not find a methodology which we believe represents well the uncertainty associated with our estimate. We therefore chose to use an uncertainty of about one and a half orders of magnitude for both the number of bacteriophages in the marine environments and in the marine deep subsurface. We hope further studies could come up with a better methodology for assessing the uncertainty of the estimate of the total number of bacteriophages in those environments.
We combine the uncertainties associated with the estimates of the total number of viruses in each environment, and arrive at a total uncertainty of about 13-fold for the total number of viruses globally (link to full calculation).

To quantify the uncertainty associated with the size range of viruses, we use data from Brum et al. (278) on marine viruses. The uncertainty of the average diameter of marine viruses is less than 2-fold (link to full calculation). This estimate doesn't take into consideration the fact that viruses from other environments might have different size ranges. Nevertheless, bacteriophages isolated from deep terrestrial deep subsurface (280) are also in the same range. We propagate the uncertainty in the diameter of bacteriophages into our estimate for the carbon content of a single bacteriophage, which results in an uncertainty of ≈2-fold (link to full calculation). We combine this uncertainty in our estimate of the radius of bacteriophages with the uncertainty the parameters of the biophysical model introduce into the estimate of the carbon content of bacteriophages. The uncertainty of the parameters of the model are detailed in Jover et al. (279). Overall, we project an uncertainty of ≈2-fold associated with our estimate of the carbon content of a single bacteriophage (link to full calculation). We note a caveat in regard to the carbon content of viruses due to the fact that we estimate the same carbon content for both archaeal and bacterial viruses, even though archaeal viruses are highly diverse (281). Nevertheless, the relative biomass of bacteria is higher than that of archaea, and thus it is likely that bacterial viruses are more abundant that archaeal viruses.

Combining the uncertainty of the number of cells in the deep terrestrial deep subsurface with the uncertainty of the ratio between viruses and bacteria, and also the uncertainty of the size of viruses and their carbon content, we project a total uncertainty of about ≈15-fold (link to full

calculation). Due to the scarcity of data on the different parameters used to estimate the total biomass of phages, we use a higher uncertainty projection of ≈20-fold.

Other Animal Phyla

In this work, we survey the biomass of key kingdoms in the tree of life. For all kingdoms except animals, estimates on the biomass of those kingdoms is made in a top-down manner, where the biomass of all the different taxa making up the kingdom are considered together. This is largely due to the morphological similarity between organisms in a kingdom, or because of technical capabilities that allow to estimate the entire biomass of the kingdom, like the quantification of chlorophyll. For animals, however, the large diversity of life forms and the distribution of animals both on land and at sea doesn't allow a top-down approach to estimation of animal biomass. Therefore, for estimating the biomass of animals, we use a bottom-up approach, which estimates the biomass of key phyla constituting the animal kingdom, and the sum of the biomass of those phyla represents our estimate of the total biomass of animals. For some phyla, such as echinoderms and bryozoans, we could not find explicit data in the literature to support a full-fledged biomass estimate. We try to estimate bounds for the possible biomass contribution for these remaining phyla. Most of the remaining phyla not covered by the analysis are aquatic, and for the large part, benthic organisms. The only major phyla which are terrestrial are tardigrades (water bears) and rotifers. The available data suggest that the biomass contribution of those taxa is negligible when considering the entire animal biomass (146, 282). Therefore, putting a bound on the global benthic biomass could serve as an upper estimate of their biomass which we now proceed to infer.

Benthic phyla biomass

As part of the Census of Marine Life (283), a decade long international scientific effort to assess the diversity, distribution and abundance of marine life, Wei et al. (87) reported an estimate of ≈0.1 Gt C for the global benthic biomass, with bacteria contributing ≈30% of this biomass. Thus, we are left with ≈0.06 Gt C of eukaryotic benthic biomass. There are, however, several caveats associated with this estimate, as the estimate is related mainly to muddy surfaces in the ocean, and doesn't take into account possible biomass hotspots such as seamounts, submarine canyons and trenches, hydrothermal vents and cold seeps. Additionally, it doesn't take into account contribution from benthic foraminifera, which can sometimes account for more than 50% of eukaryote biomass in specific sites (284).
We try to estimate the contribution of seamounts and canyons from local biomass density samplings. For seamounts, high megafaunal biomass densities were measured in the southwest Pacific seamounts (285). About ≈70 g C m^{-2} of megafaunal biomass were estimated, as opposed to ≈20 g C m^{-2} megafaunal biomass in the so-called slopes with comparable depth which are not associated with seamounts. Multiplying this biomass density with the total seamount area of ≈9×10^{12} m^2 (212), we get an estimate of ≈0.6 Gt C. This estimate only considers megafaunal biomass. On the other hand, this reference is based on comparison to non-seamount area and for those it uses a much higher value than a previous report (87). It is thus unclear if this is an under- or over-estimate. For submarine canyons, megafaunal biomass density was estimated at 89 g C m^{-2} in the Kaikoura Canyon on the eastern New Zealand margin (286). This reported biomass is

≈100 fold larger than values reported for general deep-sea environment (87). Such high biomass density was estimated to be present in ≈15% of all submarine canyons (286). The total submarine canyon area is roughly $\approx 4\times10^{12}$ m^2 (212), so the total estimate for megafaunal biomass in canyons is ≈0.06 Gt C. For trenches, benthic biomass was estimated to be on the order of ≈10 g C m^2 (287). The global area of oceanic trenches is $\approx 2\times10^{12}$ m^2 (212), so the total trench benthic biomass is estimated at ≈0.02 Gt C. For benthic foraminifera, we use a geometric mean of previously reported biomass densities (284), which is roughly ≈0.16 g C m^{-2}, and multiply this density with the global surface area of the sea, which is $\approx 3.6\times10^{14}$ m^2. This yields a global biomass estimate of ≈0.05 Gt C.

For hydrothermal vents, biomass densities on the order of ≈1000 g C m^{-2} were reported in several locations (288). A current map of active hydrothermal vents estimates ≈500 active sites (289). Assuming a characteristic area of a vent of $\approx 10^6$ m^2 (290) we estimate a global biomass of ≈0.0005 Gt C.

All in all, we estimate the global benthic biomass to be smaller than 1 Gt C. It is important to remember, however, that most of the studies surveying biomass hotspots are focused on megafauna or macrofauna, so data on other animal groups is underrepresented. Although estimates for each contribution to the total benthic biomass are coarse, we believe 1 Gt C is a probable upper bound for the benthic biomass. From this upper bound estimate, we can infer that benthic biomass will not change dramatically the distribution of biomass presented in this work. Because this is only an upper bound and the real value might be much lower this benthic biomass was not portrayed in the figures and table in the main text of the work.

Pre-human biomass

The impact of humans on the distribution of biomass on Earth has begun long before present times. For example, a large extinction event, the Quaternary Megafauna Extinction, which occurred 50,000-7,000 years ago, is at least partially explained by human hunting and habitat alteration (175). This extinction claimed a half of megafaunal species (defined as species weighing over 100 lb, i.e. 44 kg). We estimate the total wild mammal biomass before the Quaternary Megafauna Extinction event and compare it with present day values. The overall biomass of wild mammals is dominated by large mammals, i.e megafauna (173). Moreover, the estimated megafauna biomass before the Quaternary Megafauna Extinction event is dominated by mammals (175). Therefore, the biomass values of megafauna given below are a good approximation to to the overall biomass of wild mammals. The global biomass of terrestrial megafauna before the Quaternary Megafauna Extinction event is estimated to have been around ≈0.02 Gt C (175). The global biomass of marine wild mammals before human exploitation is estimated at ≈0.02 Gt C (178). The total wild mammal biomass before the Quaternary Megafauna Extinction event of ≈0.04 Gt C is about 6-fold higher than the ≈0.007 Gt C of extant wild mammal biomass. We cannot currently derive the uncertainty associated with the change in wild mammal biomass before and after human civilization, as we do not have a projection for the uncertainty associated with the pre-human wild land mammal biomass. We hope further research will help better assess the certainty of estimates regarding the biomass of pre-human wild land mammals. As for the pre-human marine wild mammal biomass, Christensen (178) reports a 95% confidence interval of ≈1.3-fold. Regarding the change in wild marine mammal biomass,

Christensen reports a range of 2.4-fold to 7-fold. At the same time that the biomass of wild megafauna collapsed, the biomass of humans gradually increased over the same period. Since the industrial revolution we have witnessed an exponential increase in human population, as well as a rapid increase in the domesticated livestock biomass. Today, the biomass of livestock (≈0.1 Gt) is an order of magnitude larger than that of all the terrestrial wild megafauna before the Quaternary Megafauna Extinction. Even the biomass of humans alone (≈0.05 Gt) is around twice the size of the biomass of all wild megafauna before the Quaternary Megafauna Extinction event. In addition to the influence humans had on mammal biomass, the expansion of human population in recent centuries has dramatically changed many habitats. One prominent impact humans have had on wild populations was caused by fishing. In order to estimate the relation between current fish biomass and pre-human fish biomass, we rely on models of fishery ecology (291), in conjunction with data on the current state of fisheries. A recent study (292), surveyed the status of fisheries which account for ≈80% of the global fish catch, and estimated that fish biomass is ≈120% of the biomass which will ensure maximum sustainable yields (see Table S12 in Costello et al). Costello et al. puts current fisheries biomass at ≈0.84 Gt wet weight (292), or ≈0.13 Gt C, which is on par with previous estimates for the non-mesopelagic fish biomass (161). Using a database of published landings data and stock assessment biomass estimates, Thorson et al. (293) estimate that the biomass of fish at the maximum sustainable yield represent ≈40% of the biomass the population would have reached in case of no fishing. From these studies, we estimate that before human civilization started fishing, the biomass of fisheries was ≈2-fold larger than current values (1/0.4×1/1.2). If we take estimates for present day fisheries biomass at ≈0.13 Gt C, this translated to ≈0.27 Gt C of fisheries before human civilization (link to full calculation). Assuming that the biomass of mesopelagic fish was not influenced significantly by fishing, we estimate the total global biomass of fish before human civilization across all depths at ≈0.8 Gt C. Many of the impacts humans had on the biomass of wild populations are hard to quantify, as accurate estimates of the standing stocks of fauna before human development are missing. For plants, however, some estimates are available. The biomass of plants is dominated by trees. Estimates put the global biomass of trees before human civilization at around twice its current value (294). As plants are the dominant fraction of global biomass, this means that humans have reduced the total biomass of the biosphere to about half of its pre-human value.

Microbiomes

The microbiome, the set of commensal, symbiotic and pathogenic microbes that reside in the body of a host organism, has gained increasing attention due to its role in health and disease. Here we focus on animal microbiome and quantify their contribution to global biomass of prokaryotes. In their pioneering study, Whitman et al. (25) estimated the contribution of prokaryotes residing in humans, livestock and termites. They showed the contribution from those sources to the global pool of prokaryotes in negligible (about one millionth of the total number of prokaryotes). A recent study (295) found an allometric relation between the number of prokaryotes in the microbiome of various animals and the mass of the host. They find that the number of microbes per animal scales like $7.86\times10^{8}\times M^{1.07}$, where M is the wet weight mass of the animal in units of gram wet weight. This power law fits the data impressively well (R^2=0.94). The fact that the power of the exponent is very close to one suggests that this relation can also be expressed in

linear terms. Indeed, a linear fit of the data yields a relation of $3.4\times10^9\times M$, still with a high correlation coefficient ($R^2=0.8$). The authors then used these relations, along with an estimate of the total mass of animals to produce a global estimate of $\approx2\times10^{25}$ microbial cells associated with animals. The estimates for the total mass of animals the authors used are based on previous studies, and one of our main results is an updated account of the total mass of animals. We thus apply the allometric relation found in their study to the updated total animal biomass which we estimate of 3 Gt C. As we estimate the biomass of animals in weight of carbon, and the allometric relation refers to wet weight, we translate our estimate to wet weight by dividing it by 0.15, the ratio of carbon to wet weight assuming ≈70% water content and ≈50% carbon content out of dry weight. This yields an estimate of ≈20 Gt wet weight of animals. Applying the linear relation found in their paper (295), we estimate a total of $\sim10^{26}$ microbial cells associated with animals. This is indeed a tiny fraction of the total estimated number of prokaryotic cells ($\approx10^{30}$). Assuming an *E. coli* like carbon content per prokaryotic cell of ≈150 fg C, the total mass of animal associated microbes will be ~0.01 Gt C.

Inland water

In our study we concentrate on estimating the biomass in marine, terrestrial or deep subsurface environments. Biomass is also present in inland waters such as lakes or rivers, but its contribution to the global biomass of each taxon is negligible due to the small area which inland waters cover. Recent reports estimate the total area covered by lakes at $\approx3\times10^6$ km^2 (296). In order to estimate the maximum contribution of the biomass inland waters could have, we multiply this area by the maximum biomass concentration per area reported for phytoplankton, zooplankton and fish across 24 lakes worldwide (297). Summing up the maximum possible contribution from each group yields a biomass concentration of ≈50 g C m^{-2}. Multiplying this concentration by the global area of lakes constrains the total biomass present in inland waters below ≈0.15 Gt C, which is still negligible compared to contributions from other environments.

Abundance estimates

General

We provide detailed derivation for the estimate of the number of individuals in each taxon. For some groups, the estimate of the number of total individuals has already been derived as a part of estimating the global biomass of the group. For other groups, additional steps were needed in order to estimate their total number of individuals. Here, we discuss the groups for which estimates are already available. In the sections below, we give explain in more detail the estimates of specific groups for which additional work was needed. For plants, we estimate the total number of trees, based on a recent study (294). Including all plant species will definitely increase the estimate dramatically, but due to the high diversity of species and characteristic sizes of different plant species, it is very difficult to estimate the total number of plants in the biosphere. For annelids and nematodes, we use our biomass estimates and divide them by the mean individual body mass that was used by Fierer et al. (144) to estimate their biomass densities. For humans, we rely on the UN estimate of the global population. For livestock we rely

on the FAOStat database (http://faostat3.fao.org/; domain: Production/Live animals). For birds, we rely on estimates for the global population of birds (180).

Archaea and Bacteria (Prokaryotes)

For marine prokaryotes, as well as marine and terrestrial subsurface prokaryotes, in order to estimate the global biomass, we estimate the total number of cells, and multiply it by the average cell mass. Thus, for those groups an estimate of the total number of cells is already contained in the estimate of their global biomass (1.2×10^{29}, 4×10^{29} and 2×10^{30} respectively). We calculate the number of bacteria and archaea out of the total population of prokaryotes using our estimates for their relative fraction out of the biomass of prokaryotes, assuming the size of bacteria is not significantly different than the size of archaea. For soil prokaryotes, however, biomass estimates are based on measurement of total microbial biomass. Therefore, we take the total biomass estimate and divide it by a characteristic cell mass. We use the same masses of bacteria from other environments of ≈30 fg C $cell^{-1}$ (298). As we estimate about 8 Gt C of soil prokaryote biomass, we estimate about $\approx3\times10^{29}$ prokaryotic cells in the soil. From the total number of prokaryotes in the soil, we estimate the number of bacteria and archaea by using their estimated fractions out of the total biomass of prokaryotes. Even though soil archaea are smaller than soil bacteria, and thus we probably underestimate the number of soil archaea, this will not affect the estimate for the total amount of archaea as the contribution from soil archaea is small.

Fungi

When estimating the biomass of fungi globally, we use estimates for the total microbial biomass in the soil, as well as estimates for the fraction of soil microbial biomass which is fungal. In addition, we use estimates of the biomass fraction of mycorrhiza out of fine-root biomass. As these estimates do not rely on estimating the total number of fungal cells, we derive our estimate for the total number of fungal cells by dividing our biomass estimate by an average carbon content per fungal cell. We very roughly estimate the volume of fungal cells to be ≈100 μm^3 (299), and thus we estimate a carbon content of a cell to be ≈15 pg C $cell^{-1}$. Dividing our estimate for the total fungal biomass of ≈12 Gt C by the average carbon content per fungal cell, we get $\sim10^{27}$ fungal cells globally (link to full calculation).

Molluscs

The estimate of the biomass of pteropods is based on a database which contains both abundance data and species description. This data was converted to biomass based on the average mass of each species. In order to estimate the number of individual pteropods, we use the data from the MAREDAT database (300) crudely divide the average biomass density of pteropods by the average population density of pteropods. This gives us an average mass of individual pteropods. We get an average individual mass of ≈0.3 mg C per individual pteropod. We estimate the total biomass of pteropods to be ≈0.15 Gt C. Dividing the global biomass of pteropods by the average mass of an individual pteropod brings the total number of individual pteropods to be around 5×10^{17} (link to full calculation).

Fish

In order to estimate the total number of fish, we rely on our estimate for the total biomass of fish of ≈0.5 Gt C and estimate the mean mass of an individual mesopelagic fish, as mesopelagic fish are the main contributor to the total fish biomass. In order to estimate the mean mass of an individual mesopelagic fish, we use empirical allometric relations between fish length and mass, along with ranges for the lengths of different mesopelagic fish species (301). From the available data, we estimate an average individual mass of ≈0.5 g C per fish (link to full calculation). This brings our estimate of the total number of mesopelagic fish to $\approx 10^{15}$ fish.

Arthropods

For terrestrial arthropods, we rely on the estimate made originally by Williams (195) of 10^{17}-10^{19}. We use the geometric mean of 10^{18} as our best estimate. For marine arthropods, we estimate the total number of individuals to be dominated by copepods, as they are usually on the lower end of the marine arthropod body size while still representing a large fraction of the marine arthropod mass. The estimate of the biomass of copepods is based on the estimate of mesozooplankton biomass, as it is defined to contain all copepod biomass (206). Copepods vary in size, some are considered microzooplankton, while others are considered either mesozooplankton or macrozooplankton. As all copepod biomass was concentrated under the mesozooplankton group, we do not know the partitioning of biomass between size classes, so we assume they are all mesozooplankton. The values reported in the literature (302, 303), range from 0.15 µg C to 100 µg C (we note this is based on a rather limited number of sources). We take the geometric mean of this range, which is ≈4 µg C per individual. The total biomass of mesozooplankton is ≈0.6 Gt C, so by dividing the total biomass by the estimated mass of an individual copepod we get $\sim 10^{20}$ copepods (link to full calculation).

Cnidaria

The biomass of cnidarians was estimated by using a database of densities of individuals, along with estimates of the mean mass of each species of cnidarian. In order to estimate the total number of cnidarians, we use several measures of the characteristic mass of an individual cnidarian. The median mass of an individual cnidarian is ≈0.5 mg C (211). The total estimate of planktonic cnidaria is ≈0.05 Gt C. This means that the estimate for the total abundance of cnidaria is $\sim 10^{17}$. If instead of using the median value we use the average mass value we get ≈7 g C per individual (211), which translates into $\approx 7 \times 10^{15}$ individual cnidarians. The geometric mean of both estimates is $\approx 2 \times 10^{16}$ individuals. If we decide to take the average density of individuals from the database and extrapolate it globally to the top 500 m we get $\approx 10^{16}$ individuals. Our best estimate is the geometric mean of these three methods, which is $\approx 2 \times 10^{16}$ (link to full calculation).

Protists

The biomass of protists is composed of 0.4 Gt C nano-pico protists, 0.5 Gt C microzooplankton, 0.5 Gt C Diatoms, 0.5 Gt C Phaeocystis. As the nano-pico are the smallest and have a comparable overall biomass to the other class sizes, they will dominate the abundance of protists. The

diameter range of pico-nanoplankton is 0.8-5 μm. We use the geometric mean of the radius range, which is ≈1 μm. This means that the mean cell volume is ≈4 μm^3. Using the conversion equation between biovolume and carbon content reported in Pernice et al. (132), we estimate an average single protist will have $\approx 0.8 \times 10^{-12}$ g C $cell^{-1}$ (link to full calculation). Dividing the total 0.4 Gt C by the carbon content of a single cell we estimate $\sim 10^{27}$ protist cells. For terrestrial protists, we estimate the total number of individuals in the process of estimating their total biomass. We estimate a total of $\approx 3 \times 10^{25}$ terrestrial protists, which is much lower than our estimate of $\sim 10^{27}$ marine protists.

Usage of various estimators of the mean

As we note in the methods section of the main text, in order to generate global estimates based on local sampling, we calculate average values from all available peer-reviewed literature sources we could find and extrapolate from them to the global scale. There are two distinct types of averages we calculate. The first is calculating a characteristic value from different independent measurements, for example, calculating the characteristic carbon content of bacteria, from different studies which report different carbon content values. In this case, we consider the reported values from different studies to be independent samples from the distribution of the carbon content of bacteria. Deviations between values reported in different studies are often the result of different measurement methodologies or biases rather than actual differences in the populations sampled. In such cases we use the geometric mean of the values from the different studies as our best estimate, because we assume, for lack of better knowledge, that the measurement error, in this specific example the measurement error of the carbon content of bacteria, is multiplicative rather than additive. The distribution is thus better approximated as log-normal than normal and the geometric mean gives the most probable value.

The second kind of average values we calculate is when we use different samples of the population density of organisms in the environment and calculate the global average population density. In this case, even if the samples are log-normally distributed, the arithmetic mean gives the average population density, as it allows, for example, especially high values to shift the average. Yet using the arithmetic mean of the samples has the disadvantage of being more susceptible to biases in oversampling singular locations with high values which are often locations of research interest, such as blooms in the case of phytoplankton. Using the arithmetic mean in this case might lead to a large overestimate. On the other hand, using as an alternative the geometric mean of the samples might underestimate the true average population density as it will reduce the effect of biologically relevant high population densities, which are the result of heterogeneous distribution of organisms across the globe. As each method has advantages and disadvantage, we chose throughout our analysis to calculate the average population density based both on the arithmetic mean and based on the geometric mean. We treat these two average population densities as two estimates of the actual average population density. As our best estimate of the average population density we use the geometric mean of these two estimates (i.e. the geometric mean of the arithmetic and geometric means). This approach, while not standard, was chosen as it increases our robustness to the possible biases discussed above. A most standard approach could be to use the median over the reported values, though it will also reduce the effect of locations with especially high population density. For comparison, we analyzed our results also

using this alternative approach of taking the median and found it does not change our final results beyond the multiplicative standard deviation of our estimates, thus showing that our estimates are robust well within the uncertainty we report.

Supplementary References

1. Pan Y, et al. (2011) A large and persistent carbon sink in the world's forests. *Science* 333(6045):988–993.

2. MacDicken KG (2015) Global Forest Resources Assessment 2015: What, why and how? *For Ecol Manage* 352:3–8.

3. Saatchi SS, et al. (2011) Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc Natl Acad Sci U S A* 108(24):9899–9904.

4. Baccini A, et al. (2012) Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nat Clim Chang* 2(3):nclimate1354.

5. Thurner M, Beer C, Santoro - Global Ecology and … M, 2014 (2014) Carbon stock and density of northern boreal and temperate forests. *Wiley Online Library*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/geb.12125/full.

6. Saugier B, Roy J, Mooney HA (2001) Estimations of Global Terrestrial Productivity. *Terrestrial Global Productivity* (Elsevier), pp 543–557.

7. Ajtay GL (1979) Terrestrial primary production and phytomass. *The global carbon cycle, SCOPE 13*:129–181.

8. Ruesch A, Gibbs H (2008) New IPCC Tier-1 Global Biomass Carbon Map for the Year 2000 (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory).

9. Erb K-H, et al. (2017) Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature*. doi:10.1038/nature25138.

10. Smith R (2001) Global Forest Resources Assessment 2000 Main Report. *Food and Agriculture Organization: Rome, Italy*. Available at: http://library.nust.ac.zw/gsdl/collect/digitisa/import/Digital%20Libraries/lesson0301.pdf.

11. Hutchison J, et al. (2013) Predicting Global Patterns in Mangrove Forest Biomass. *Conservation Letters* 7(3):233–240.

12. Donato DC, et al. (2011) Mangroves among the most carbon-rich forests in the tropics. *Nat Geosci* 4(5):293–297.

13. Fourqurean JW, et al. (2012) Seagrass ecosystems as a globally significant carbon stock. *Nat Geosci* 5(7):505–509.

14. Elbert W, et al. (2012) Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. *Nat Geosci* 5(7):459–462.

15. Edwards D, Dianne E, Lesley C, Raven JA (2015) Could land-based early photosynthesizing ecosystems have bioengineered the planet in mid-Palaeozoic times? *Palaeontology* 58(5):803–837.

16. De Vooys C (1979) SCOPE 13--The Global Carbon Cycle. Available at: http://www.scopenvironment.org/downloadpubs/scope13/chapter10.html.

17. Charpy Roubaud C, Sournia A (1990) The comparative estimation of phytoplanktonic, microphytobenthic and macrophytobentic primary production in the oceans. *Mar Microb Food Webs* 4(1):31–57.

18. Smith SV (1981) Marine macrophytes as a global carbon sink. *Science* 211(4484):838–840.

19. Poorter H, et al. (2012) Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control. *New Phytol* 193(1):30–50.

20. Erb K-H, et al. (2016) Biomass turnover time in terrestrial ecosystems halved by land use. *Nat Geosci* 9(9):ngeo2782.

21. Jackson RB, et al. (1996) A global analysis of root distributions for terrestrial biomes. *Oecologia* 108(3):389–411.

22. Asner GP, Scurlock JMO, A. Hicke J (2003) Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Glob Ecol Biogeogr* 12(3):191–205.

23. Schlesinger WH, Bernhardt ES (2013) The Global Carbon Cycle. *Biogeochemistry*, pp 419–444.

24. Wright IJ, et al. (2004) The worldwide leaf economics spectrum. *Nature* 428(6985):821–827.

25. Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 95(12):6578–6583.

26. Arístegui J, Gasol JM, Duarte CM, Herndld GJ (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* 54(5):1501–1529.

27. Buitenhuis ET, et al. (2012) Global distributions of picoheterothrophs (Bacteria and Archaea) abundance and biomass-Gridded data product (NetCDF)-Contribution to the MAREDAT World Ocean Atlas of Plankton Functional Types.

28. Lloyd KG, May MK, Kevorkian RT, Steen AD (2013) Meta-analysis of quantification methods shows that archaea and bacteria have similar abundances in the subseafloor. *Appl Environ Microbiol* 79(24):7790–7799.

29. Arístegui J, Javier A, Gasol JM, Duarte CM, Herndld GJ (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* 54(5):1501–1529.

30. Buitenhuis E, et al. (2012) Picoheterotroph (Bacteria and Archaea) biomass distribution in the global ocean. *Earth System Science Data* 4(1):101–106.

31. Lee S, Fuhrman JA (1987) Relationships between Biovolume and Biomass of Naturally Derived Marine Bacterioplankton. *Appl Environ Microbiol* 53(6):1298–1303.

32. Ducklow HW, Carlson CA (1992) Oceanic Bacterial Production. *Advances in Microbial Ecology*, Advances in Microbial Ecology. (Springer US), pp 113–181.

33. Gundersen K, Heldal M, Norland S, Purdie DA, Knap AH (2002) Elemental C, N, and P cell content of individual bacteria collected at the Bermuda Atlantic Time-series Study (BATS) site. *Limnol Oceanogr* 47(5):1525–1530.

34. Carlson CA, Bates NR, Ducklow HW, Hansell DA (1999) Estimation of bacterial respiration and growth efficiency in the Ross Sea, Antarctica. *Aquat Microb Ecol* 19:229–244.

35. Fukuda R, Ogawa H, Nagata T, Koike I I (1998) Direct determination of carbon and nitrogen contents of natural bacterial assemblages in marine environments. *Appl Environ Microbiol* 64(9):3352–3358.

36. Simon M, Grossart HP, Schweitzer B, Ploug H (2002) Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat Microb Ecol* 28:175–211.

37. Alldredge AL, Gotschalk CC (1990) The relative contribution of marine snow of different origins to biological processes in coastal waters. *Cont Shelf Res* 10(1):41–58.

38. Turley CM, Stutt ED (2000) Depth-related cell-specific bacterial leucine incorporation rates on particles and its biogeochemical significance in the Northwest Mediterranean. *Limnol Oceanogr* 45(2):419–425.

39. Cho BC, Azam F (1988) Major role of bacteria in biogeochemical fluxes in the ocean's interior. *Nature* 332:441.

40. Alldredge AL, Cole JJ, Caron DA (1986) Production of heterotrophic bacteria inhabiting macroscopic organic aggregates (marine snow) from surface waters1. *Limnol Oceanogr* 31(1):68–78.

41. Turley CM, Mackie PJ (1994) Biogeochemical significance of attached and free-living bacteria and the flux of particles in the NE Atlantic Ocean. *Mar Ecol Prog Ser* 115(1/2):191–203.

42. Simon M, Alldredge AL, Azam F (1990) Bacterial carbon dynamics on marine snow. *Mar Ecol Prog Ser* 65(3):205–211.

43. Herndl GJ, Peduzzi P (1988) The Ecology of Amorphous Aggregations (Marine Snow) in the Northern Adriatic Sea: *Mar Ecol* 9(1):79–90.

44. Thiele S, Fuchs BM, Amann R, Iversen MH (2015) Colonization in the photic zone and subsequent changes during sinking determine bacterial community composition in marine snow. *Appl Environ Microbiol* 81(4):1463–1471.

45. Bochdansky AB, van Aken HM, Herndl GJ (2010) Role of macroscopic particles in deep-sea oxygen consumption. *Proc Natl Acad Sci U S A* 107(18):8287–8291.

46. Gardner WD, Walsh ID (1990) Distribution of macroaggregates and fine-grained particles across a continental margin and their potential role in fluxes. *Deep Sea Res A* 37(3):401–411.

47. Honjo S, Doherty KW, Agrawal YC, Asper VL (1984) Direct optical assessment of large amorphous aggregates (marine snow) in the deep ocean. *Deep Sea Res A* 31(1):67–76.

48. Busch K, et al. (2017) Bacterial Colonization and Vertical Distribution of Marine Gel Particles (TEP and CSP) in the Arctic Fram Strait. *Frontiers in Marine Science* 4:166.

49. Passow U, Alldredge AL (1994) Distribution, size and bacterial colonization of transparent exopolymer particles (TEP) in the ocean. *Mar Ecol Prog Ser* 113(1/2):185–198.

50. Bochdansky AB, Clouse MA, Herndl GJ (2017) Eukaryotic microbes, principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. *ISME J* 11(2):362–373.

51. Karner MB, DeLong EF, Karl DM (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409(6819):507–510.

52. Herndl GJ, et al. (2005) Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* 71(5):2303–2309.

53. Teira E, Lebaron P, Van Aken H (2006) Distribution and activity of Bacteria and Archaea in the deep water masses of the North Atlantic. *Limnology and*. Available at: http://onlinelibrary.wiley.com/doi/10.4319/lo.2006.51.5.2131/full.

54. Sunagawa S, et al. (2015) Structure and function of the global ocean microbiome. *Science* 348(6237):1261359.

55. Salazar G, et al. (2016) Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J* 10(3):596–608.

56. Sun D-L, Jiang X, Wu QL, Zhou N-Y (2013) Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl Environ Microbiol* 79(19):5962–5969.

57. Kirchman DL, Elifantz H, Dittel AI, Malmstrom RR, Cottrell MT (2007) Standing stocks and activity of Archaea and Bacteria in the western Arctic Ocean. *Limnol Oceanogr* 52(2):495–507.

58. Bates ST, et al. (2011) Examining the global distribution of dominant archaeal populations in soil. *ISME J* 5(5):908–917.

59. Hong J-K, Cho J-C (2015) Environmental Variables Shaping the Ecological Niche of Thaumarchaeota in Soil: Direct and Indirect Causal Effects. *PLoS One* 10(8):e0133763.

60. Pozdnyakov LA, Stepanov AL, Manucharova NA (2011) Anaerobic methane oxidation in soils and water ecosystems. *Moscow Univ Soil Sci Bull* 66(1):24–31.

61. Lysak LV, Lapygina EV, Kadulin MS, Konova IA (2014) Number, viability, and diversity of the filterable forms of prokaryotes in sphagnous high-moor peat. *Biol Bull Russ Acad Sci* 41(3):228–232.

62. Dedysh SN, Pankratov TA, Belova SE, Kulichevskaya IS, Liesack W (2006) Phylogenetic analysis and in situ identification of bacteria community composition in an acidic Sphagnum peat bog. *Appl Environ Microbiol* 72(3):2110–2117.

63. Manucharova NA (2009) The microbial destruction of chitin, pectin, and cellulose in soils. *Eurasian Soil Sci* 42(13):1526.

64. Lysak LV, Lapygina EV, Konova IA, Zvyagintsev DG (2010) Quantity and taxonomic composition of ultramicrobacteria in soils. *Microbiology* 79(3):408–412.

65. Semenov MV, Manucharova NA, Stepanov AL (2016) Distribution of Metabolically Active Prokaryotes (Archaea and Bacteria) throughout the Profiles of Chernozem and Brown Semidesert Soil. *Eurasian Soil Sci* 49(2):217–225.

66. Kobabe S, Wagner D, Pfeiffer E-M (2004) Characterisation of microbial community composition of a Siberian tundra soil by fluorescence in situ hybridisation. *FEMS Microbiol Ecol* 50(1):13–23.

67. Šimek M, et al. (2014) The microbial communities and potential greenhouse gas production in boreal acid sulphate, non-acid sulphate, and reedy sulphidic soils. *Sci Total Environ* 466-467:663–672.

68. Barra Caracciolo A, et al. (2015) Changes in microbial community structure and functioning of a semiarid soil due to the use of anaerobic digestate derived composts and rosemary plants. *Geoderma* 245-246(Supplement C):89–97.

69. Sheibani S, et al. (2013) Soil bacteria and archaea found in long-term corn (Zea mays L.) agroecosystems in Quebec, Canada. *Can J Soil Sci* 93(1):45–57.

70. Eickhorst T, Tippkötter R (2008) Improved detection of soil microorganisms using fluorescence in situ hybridization (FISH) and catalyzed reporter deposition (CARD-FISH). *Soil Biol Biochem* 40(7):1883–1891.

71. Schmidt H, Eickhorst T (2013) Spatio-temporal variability of microbial abundance and community structure in the puddled layer of a paddy soil cultivated with wetland rice (Oryza sativa L.). *Appl Soil Ecol* 72(Supplement C):93–102.

72. Ushio M, Makoto K, Klaminder J, Nakano S-I (2013) CARD-FISH analysis of prokaryotic community composition and abundance along small-scale vegetation gradients in a dry arctic tundra ecosystem. *Soil Biol Biochem* 64(Supplement C):147–154.

73. Orcutt BN, Sylvan JB, Knab NJ, Edwards KJ (2011) Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol Mol Biol Rev* 75(2):361–422.

74. Schrenk MO, Huber JA, Edwards KJ (2010) Microbial provinces in the subseafloor. *Ann Rev Mar Sci* 2:279–304.

75. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc Natl Acad Sci U S A* 109(40):16213–16216.

76. Parkes RJ, et al. (2014) A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. *Mar Geol* 352:409–425.

77. Parkes RJ, et al. (1994) Deep bacterial biosphere in Pacific Ocean sediments. *Nature* 371(6496):410–413.

78. Fry JC (1990) 2 Direct Methods and Biomass Estimation. *Methods in Microbiology*, eds Grigorova R, Norris JR (Academic Press), pp 41–85.

79. Lipp JS, Morono Y, Inagaki F, Hinrichs K-U (2008) Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* 454(7207):991–994.

80. Biddle JF, et al. (2006) Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *Proc Natl Acad Sci U S A* 103(10):3846–3851.

81. Simon M, Azam F (1989) Protein content and protein synthesis rates of planktonic marine bacteria. *Mar Ecol Prog Ser* 51:201–213.

82. Braun S, et al. (2016) Size and Carbon Content of Sub-seafloor Microbial Cells at Landsort Deep, Baltic Sea. *Front Microbiol* 7:1375.

83. Ingraham JL, Maaløe O, Neidhardt FC (1983) *Growth of the bacterial cell* (Sinauer Associates).

84. Xie S, Lipp JS, Wegener G, Ferdelman TG, Hinrichs K-U (2013) Turnover of microbial lipids in the deep biosphere and growth of benthic archaeal populations. *Proc Natl Acad Sci U S A* 110(15):6010–6014.

85. Hoehler TM, Jørgensen BB (2013) Microbial life under extreme energy limitation. *Nat Rev Microbiol* 11(2):83–94.

86. Kubo K, et al. (2012) Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *ISME J* 6(10):1949–1965.

87. Wei C-L, et al. (2010) Global patterns and predictions of seafloor biomass using random forests. *PLoS One* 5(12):e15323.

88. Danovaro R, Corinaldesi C, Rastelli E, Dell'Anno A (2015) Towards a better quantitative assessment of the relevance of deep-sea viruses, Bacteria and Archaea in the functioning of the ocean seafloor. *Aquat Microb Ecol* 75(1):81–90.

89. Rex MA, et al. (2006) Global bathymetric patterns of standing stock and body size in the deep-sea benthos. *Mar Ecol Prog Ser* 317:1–8.

90. Heberling C, Lowell RP, Liu L, Fisk MR (2010) Extent of the microbial biosphere in the oceanic crust. *Geochem Geophys Geosyst* 11(8):Q08003.

91. Johnson HP, Pruis MJ (2003) Fluxes of fluid and heat from the oceanic crustal reservoir. *Earth Planet Sci Lett* 216(4):565–574.

92. Jungbluth SP, Grote J, Lin H-T, Cowen JP, Rappé MS (2013) Microbial diversity within basement fluids of the sediment-buried Juan de Fuca Ridge flank. *ISME J* 7(1):161–172.

93. Jungbluth SP, Bowers RM, Lin H-T, Cowen JP, Rappé MS (2016) Novel microbial assemblages inhabiting crustal fluids within mid-ocean ridge flank subsurface basalt. *ISME J* 10(8):2033–2047.

94. Meyer JL, et al. (2016) A distinct and active bacterial community in cold oxygenated fluids circulating beneath the western flank of the Mid-Atlantic ridge. *Sci Rep* 6:22541.

95. Zhang X, Feng X, Wang F (2016) Diversity and Metabolic Potentials of Subsurface Crustal Microorganisms from the Western Flank of the Mid-Atlantic Ridge. *Front Microbiol* 7:363.

96. Mason OU, et al. (2010) First investigation of the microbiology of the deepest layer of ocean crust. *PLoS One* 5(11):e15399.

97. Harvey RW, Smith RL, George L (1984) Effect of organic contamination upon microbial distributions and heterotrophic uptake in a Cape Cod, Mass., aquifer. *Appl Environ Microbiol* 48(6):1197–1202.

98. McMahon S, Parnell J (2014) Weighing the deep continental biosphere. *FEMS Microbiol Ecol* 87(1):113–120.

99. Gleeson T, Befus KM, Jasechko S, Luijendijk E, Bayani Cardenas M (2015) The global volume and distribution of modern groundwater. *Nat Geosci* 9(2):161–167.

100. Gold T (1992) The deep, hot biosphere. *Proceedings of the National Academy of Sciences* 89(13):6045–6049.

101. Rempfert KR, et al. (2017) Geological and Geochemical Controls on Subsurface Microbial Life in the Samail Ophiolite, Oman. *Front Microbiol* 8:56.

102. Lau MCY, et al. (2016) An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *Proc Natl Acad Sci U S A* 113(49):E7927–E7936.

103. Osburn MR, LaRowe DE, Momper LM, Amend JP (2014) Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. *Front Microbiol* 5:610.

104. Simkus DN, et al. (2016) Variations in microbial carbon sources and cycling in the deep continental subsurface. *Geochim Cosmochim Acta* 173:264–283.

105. Purkamo L, et al. (2016) Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids. *Biogeosciences*. Available at: https://helda.helsinki.fi/handle/10138/165462.

106. Takai K, Moser DP, DeFlaun M, Onstott TC, Fredrickson JK (2001) Archaeal diversity in waters from deep South African gold mines. *Appl Environ Microbiol* 67(12):5750–5760.

107. Bomberg M, Lamminmäki T, Itävaara M (2016) Microbial communities and their predicted metabolic characteristics in deep fracture groundwaters of the crystalline bedrock at Olkiluoto, Finland. *Biogeosciences* 13(21):6031–6047.

108. Moser DP, et al. (2005) Desulfotomaculum and Methanobacterium spp. dominate a 4- to 5-kilometer-deep fault. *Appl Environ Microbiol* 71(12):8773–8783.

109. Wardle DA (1992) A Comparative assessment of factors which influence microbial biomass carbon and nitrogen levels in soil. *Biol Rev Camb Philos Soc* 67(3):321–358.

110. Xu X, Xiaofeng X, Thornton PE, Post WM (2012) A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. *Glob Ecol Biogeogr* 22(6):737–749.

111. Serna-Chavez HM, Noah F, van Bodegom PM (2013) Global drivers and patterns of microbial abundance in soil. *Glob Ecol Biogeogr* 22(10):1162–1172.

112. Vance ED, Brookes PC, Jenkinson DS An extraction method for measuring soil microbial biomass C - ScienceDirect. Available at: http://www.sciencedirect.com/science/article/pii/0038071787900526 [Accessed February 14, 2017].

113. Jenkinson DS (1966) Studies on the decomposition of plant material in soil. *J Soil Sci* 17(2):280–302.

114. Joergensen R, Wichern F (2008) Quantitative assessment of the fungal contribution to microbial tissue in soil. *Soil Biol Biochem* 40(12):2977–2991.

115. Perry DA, Oren R, Hart SC (2013) *Forest Ecosystems* (JHU Press).

116. Baldrian P, et al. (2013) Estimation of fungal biomass in forest litter and soil. *Fungal Ecol* 6(1):1–11.

117. Bardgett RD, van der Putten WH (2014) Belowground biodiversity and ecosystem functioning. *Nature* 515(7528):505–511.

118. Bååth E, Söderström B (1979) The Significance of Hyphal Diameter in Calculation of Fungal Biovolume. *Oikos* 33(1):11–14.

119. Terrer C, Vicca S, Hungate BA, Phillips RP, Prentice IC (2016) Mycorrhizal association as a primary control of the CO_2 fertilization effect. *Science* 353(6294):72–74.

120. Shi M, Fisher JB, Brzostek ER, Phillips RP (2016) Carbon cost of plant nitrogen acquisition: global carbon cycle impact from an improved plant nitrogen cycle in the Community Land Model. *Glob Chang Biol* 22(3):1299–1314.

121. Stögmann B, Marth A, Pernfuß B, Pöder R (2013) The architecture of Norway spruce ectomycorrhizae: three-dimensional models of cortical cells, fungal biomass, and interface for potential nutrient exchange. *Mycorrhiza* 23(6):431–445.

122. Ostonen I, Ivika O, Krista L, Katrin P (2005) Fine root biomass, production and its proportion of NPP in a fertile middle-aged Norway spruce forest: Comparison of soil core and ingrowth core methods. *For Ecol Manage* 212(1-3):264–277.

123. Yuan ZY, Chen HYH (2010) Fine Root Biomass, Production, Turnover Rates, and Nutrient Contents in Boreal Forest Ecosystems in Relation to Species, Climate, Fertility, and Stand Age: Literature Review and Meta-Analyses. *CRC Crit Rev Plant Sci* 29(4):204–221.

124. Jackson RB, Mooney HA, Schulze ED (1997) A global budget for fine root biomass, surface area, and nutrient contents. *Proc Natl Acad Sci U S A* 94(14):7362–7366.

125. Stögmann B, Marth A, Pernfuß B, Pöder R (2013) The architecture of Norway spruce ectomycorrhizae: three-dimensional models of cortical cells, fungal biomass, and interface for potential nutrient exchange. *Mycorrhiza* 23(6):431–445.

126. Treseder KK, Alison C (2006) Global Distributions of Arbuscular Mycorrhizal Fungi. *Ecosystems* 9(2):305–316.

127. Zomer RJ, et al. (2016) Global Tree Cover and Biomass Carbon on Agricultural Land: The contribution of agroforestry to global and national carbon budgets. *Sci Rep* 6:29987.

128. Wang X, et al. (2014) Distribution and diversity of planktonic fungi in the West Pacific Warm Pool. *PLoS One* 9(7):e101523.

129. Gutiérrez MH, Pantoja S, Tejos E, Quiñones RA (2011) The role of fungi in processing marine organic matter in the upwelling ecosystem off Chile. *Mar Biol* 158(1):205–219.

130. de Vargas C, et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348(6237):1261605.

131. Pernice MC, et al. (2016) Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J* 10(4):945–958.

132. Pernice MC, et al. (2015) Global abundance of planktonic heterotrophic protists in the deep ocean. *ISME J* 9(3):782–792.

133. Pachiadaki MG, Rédou V, Beaudoin DJ, Burgaud G, Edgcomb VP (2016) Fungal and Prokaryotic Activities in the Marine Subsurface Biosphere at Peru Margin and Canterbury Basin Inferred from RNA-Based Analyses and Microscopy. *Front Microbiol* 7:846.

134. Orsi WD, Edgcomb VP, Christman GD, Biddle JF (2013) Gene expression in the deep biosphere. *Nature* 499(7457):205–208.

135. Schippers A, Neretin LN (2006) Quantification of microbial communities in near-surface and deeply buried marine sediments on the Peru continental margin using real-time PCR. *Environ Microbiol* 8(7):1251–1260.

136. Sinclair JL, Ghiorse WC (1989) Distribution of aerobic bacteria, protozoa, algae, and fungi in deep subsurface sediments. *Geomicrobiol J* 7(1-2):15–31.

137. Ekendahl S, O'Neill AH, Thomsson E, Pedersen K (2003) Characterisation of yeasts isolated from deep igneous rock aquifers of the Fennoscandian Shield. *Microb Ecol* 46(4):416–428.

138. Sohlberg E, et al. (2015) Revealing the unexplored fungal communities in deep groundwater of crystalline bedrock fracture zones in Olkiluoto, Finland. *Front Microbiol* 6:573.

139. Lever MA, et al. (2015) Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiol Rev* 39(5):688–728.

140. Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* 9(2):119–130.

141. Morono Y, et al. (2011) Carbon and nitrogen assimilation in deep subseafloor microbial cells. *Proc Natl Acad Sci U S A* 108(45):18295–18300.

142. Lomstein BA, Langerhuus AT, D'Hondt S, Jørgensen BB, Spivack AJ (2012) Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* 484(7392):101–104.

143. Evans GM, Furlong JC (2003) *Environmental Biotechnology: Theory and Application* (I. K. International Pvt Ltd).

144. Fierer N, et al. (2009) Global patterns in belowground communities. *Ecol Lett* 12(11):1238–1249.

145. Lavelle P, Patrick L, Spain AV (2001) Soil Organisms. *Soil Ecology*, pp 201–356.

146. Petersen H, Luxton M (1982) A Comparative Analysis of Soil Fauna Populations and Their Role in Decomposition Processes. *Oikos* 39(3):288–388.

147. Block W, Others (1970) Micro-arthropods in some Uganda soils. *" Methods of Study in Soil Ecology". Proc. UNESCO/IBP Symp. Paris 1967.*, pp 195–202.

148. Fragoso C, et al. (1999) A survey of tropical earthworms: taxonomy, biogeography and environmental plasticity. *Earthworm management in tropical agroecosystems*:1–26.

149. Decaëns T, et al. (2004) Soil macrofaunal communities in permanent pastures derived from tropical forest or savanna. *Agric Ecosyst Environ* 103(2):301–312.

150. Jiménez JJ, et al. (1998) Earthworm communities in native savannas and man-made pastures of the Eastern Plains of Colombia. *Biol Fertil Soils* 28(1):101–110.

151. Heip CHR, et al. (2001) The role of the benthic biota in sedimentary metabolism and sediment-water exchange processes in the Goban Spur area (NE Atlantic). *Deep Sea Res Part 2 Top Stud Oceanogr* 48(14-15):3223–3243.

152. Huang Y, Yong H, Zhinan Z, Xiaoshou L (2005) Studies on the community structures of meiofauna and marine nematode at six stations in the Southern Yellow Sea, China. *J Ocean Univ China* 4(1):34–42.

153. Borgonie G, et al. (2011) Nematoda from the terrestrial deep subsurface of South Africa. *Nature* 474(7349):79–82.

154. Matthews E (1997) Global litter production, pools, and turnover times: Estimates from measurement data and regression models. *J Geophys Res D: Atmos* 102(D15):18771–18800.

155. Priha O, Smolander A (1996) Microbial biomass and activity in soil and litter under Pinus sylvestris, Picea abies and Betula pendula at originally similar field afforestation sites. *Biol Fertil Soils* 24(1):45–51.

156. Bini D, Figueiredo AF, da Silva MCP, de Figueiredo Vasconcellos RL, Elke Jurandy Bran (2013) Microbial biomass and activity in litter during the initial development of pure and mixed plantations of Eucalyptus grandis and Acacia mangium. *Rev Bras Cienc Solo* 37(1):76–85.

157. Kara O, Bolat I, Cakıroglu K, Senturk M (2014) Litter Decomposition and Microbial Biomass in Temperate Forests in Northwestern Turkey. *J Soil Sci Plant Nutr* 14(1):31–41.

158. Kurzatkowski D, et al. (2004) Litter decomposition, microbial biomass and activity of soil organisms in three agroforestry sites in central Amazonia. *Nutr Cycling Agroecosyst* 69(3):257–267.

159. Forge TA, Simard SW (2000) Trophic structure of nematode communities, microbial biomass, and nitrogen mineralization in soils of forests and clearcuts in the southern interior of British Columbia. *Can J Soil Sci* 80(3):401–410.

160. Schelfhout S, et al. (2017) Tree Species Identity Shapes Earthworm Communities. *For Trees Livelihoods* 8(3):85.

161. Wilson RW, et al. (2009) Contribution of Fish to the Marine Inorganic Carbon Cycle. *Science* 323(5912):359–362.

162. Gjøsaeter J, Kawaguchi K, Food, the United Nations AO of (1980) *A Review of the World Resources of Mesopelagic Fish* (Food and Agriculture Organization of the United Nations).

163. Lam V, Pauly D (2005) Mapping the global biomass of mesopelagic fishes. *Sea Around Us Proj Newsl* 30(4). Available at: http://www.seaaroundus.org/doc/Researcher+Publications/dpauly/PDF/2005/OtherItems/MappingGlobalBiomassMesopelagicFishes.pdf.

164. Irigoien X, et al. (2014) Large mesopelagic fishes biomass and trophic efficiency in the open ocean. *Nat Commun* 5. doi:10.1038/ncomms4271.

165. Proud R, Cox MJ, Brierley AS (2017) Biogeography of the Global Ocean's Mesopelagic Zone. *Curr Biol* 27(1):113–119.

166. Davison PC, Koslow JA (2015) Acoustic biomass estimation of mesopelagic fish: backscattering from individuals, populations, and communities. *ICES J Mar Sci*. Available at: http://icesjms.oxfordjournals.org/content/early/2015/02/19/icesjms.fsv023.short.

167. Kaartvedt S, Staby A, Aksnes DL (2012) Efficient trawl avoidance by mesopelagic fishes causes large underestimation of their biomass. *Mar Ecol Prog Ser* 456:1–6.

168. Kloser RJ, Ryan TE, Young JW, Lewis ME (2009) Acoustic observations of micronekton fish on the scale of an ocean basin: potential and challenges. *ICES J Mar Sci*. doi:10.1093/icesjms/fsp077.

169. Godø OR, Patel R, Pedersen G (2009) Diel migration and swimbladder resonance of small fish: some implications for analyses of multifrequency echo data. *ICES J Mar Sci* 66(6):1143–1148.

170. Jennings S, Collingridge K (2015) Predicting Consumer Biomass, Size-Structure, Production, Catch Potential, Responses to Fishing and Associated Uncertainties in the World's Marine Ecosystems. *PLoS One* 10(7):e0133794.

171. Dong H, Mangino J, Mcallister T, Have D (2006) *Emissions from livestock and manure management* (Intergovernmental Panel on Climate Change) Available at: http://www.citeulike.org/group/10326/article/4425199.

172. Hern WM (1999) How many times has the human population doubled? Comparisons with cancer. *Popul Environ*. Available at: http://link.springer.com/article/10.1023/A:1022153110536.

173. Smil V (2011) Harvesting the biosphere: the human impact. *Popul Dev Rev* 37(4):613–636.

174. Klein Goldewijk K, Beusen A, Van Drecht G, De Vos M (2011) The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Glob Ecol Biogeogr* 20(1):73–86.

175. Barnosky AD (2008) Megafauna biomass tradeoff as a driver of Quaternary and future extinctions. *Proceedings of the National Academy of Sciences* 105(Supplement 1):11543–11548.

176. Novosolov M, et al. (2017) Population density–range size relationship revisited. *Glob Ecol Biogeogr* 26(10):1088–1097.

177. Meiri S, Raia P, Phillimore AB (2011) Slaying dragons: limited evidence for unusual body size evolution on islands. *J Biogeogr*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2010.02390.x/full.

178. Christensen LB (2006) Reconstructing historical abundances of exploited marine mammals at the global scale. Dissertation (University of British Columbia). Available at: https://open.library.ubc.ca/cIRcle/collections/ubctheses/831/items/1.0074892.

179. Kaschner K (2004) Modelling and mapping resource overlap between marine mammals and fisheries on a global scale. Dissertation (University of British Columbia). Available at: https://open.library.ubc.ca/collections/ubctheses/831/items/1.0074881.

180. Gaston KJ, Blackburn TM (1997) How many birds are there? *Biodivers Conserv* 6(4):615–625.

181. Nee S, Sean N, Read AF, Greenwood JJD, Harvey PH (1991) The relationship between abundance and body size in British birds. *Nature* 351(6324):312–313.

182. Roll U, et al. (2017) The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nat Ecol Evol* 1(11):1677–1682.

183. Feldman A, Sabath N, Pyron RA, Mayrose I, Meiri S (2016) Body sizes and diversification rates of lizards, snakes, amphisbaenians and the tuatara. *Glob Ecol Biogeogr* 25(2):187–197.

184. O'Donnell KM, Semlitsch RD (2015) Advancing Terrestrial Salamander Population Ecology: The Central Role of Imperfect Detection. *J Herpetol* 49(4):533–540.

185. Semlitsch RD, O'Donnell KM, Thompson FR (2014) Abundance, biomass production, nutrient content, and the possible role of terrestrial salamanders in Missouri Ozark forest ecosystems. *Can J Zool* 92(12):997–1004.

186. Milanovich JR, Peterman WE (2016) Revisiting Burton and Likens (1975): Nutrient Standing Stock and Biomass of a Terrestrial Salamander in the Midwestern United States. *Copeia* 104(1):165–171.

187. Gibbons JW, et al. (2006) Remarkable amphibian biomass and abundance in an isolated wetland: implications for wetland conservation. *Conserv Biol* 20(5):1457–1465.

188. Peterman WE, Crawford JA, Semlitsch RD (2008) Productivity and significance of headwater streams: population structure and biomass of the black-bellied salamander (Desmognathus quadramaculatus). *Freshw Biol* 53(2):347–357.

189. Capps KA, Berven KA, Tiegs SD (2015) Modelling nutrient transport and transformation by pool-breeding amphibians in forested landscapes using a 21-year dataset. *Freshw Biol* 60(3):500–511.

190. Reagan DP, Waide RB (1996) *The Food Web of a Tropical Rain Forest* (University of Chicago Press).

191. Beard KH, Vogt KA, Kulmatiski A (2002) Top-down effects of a terrestrial frog on forest nutrient dynamics. *Oecologia* 133(4):583–593.

192. Petranka JW, Murray SS (2001) Effectiveness of Removal Sampling for Determining Salamander Density and Biomass: A Case Study in an Appalachian Streamside Community. *J Herpetol* 35(1):36–44.

193. Santini L, Benítez-López A, Ficetola GF, Huijbregts MAJ (2017) Length - Mass allometries in Amphibians. *Integr Zool*. doi:10.1111/1749-4877.12268.

194. Hölldobler B, Wilson EO (2009) *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies* (W. W. Norton & Company).

195. Williams CB (1960) The Range and Pattern of Insect Abundance. *Am Nat* 94(875):137–151.

196. Landenmark HKE, Forgan DH, Cockell CS (2015) An Estimate of the Total DNA in the Biosphere. *PLoS Biol* 13(6):e1002168.

197. Stork NE (1988) Insect diversity: facts, fiction and speculation*. *Biol J Linn Soc Lond* 35(4):321–337.

198. Gist CS, Crossley DA (1975) The Litter Arthropod Community in a Southern Appalachian Hardwood Forest: Numbers, Biomass and Mineral Element Content. *Am Midl Nat* 93(1):107–122.

199. Brockie RE, Moeed A (1986) Animal biomass in a New Zealand forest compared with other parts of the world. *Oecologia* 70(1):24–34.

200. Ellwood MDF, Foster WA (2004) Doubling the estimate of invertebrate biomass in a rainforest canopy. *Nature* 429(6991):549–551.

201. Dial RJ, Ellwood MDF, Turner EC, Foster WA (2006) Arthropod Abundance, Canopy Structure, and Microclimate in a Bornean Lowland Tropical Rain Forest1. *Biotropica* 38(5):643–652.

202. Stork NE (1996) Tropical forest dynamics: the faunal components. *Monographiae Biologicae*, pp 1–20.

203. Fragoso C, et al. (1999) Earthworm communities of tropical agroecosystems: origin, structure and influence of management practices. *Earthworm management in tropical agroecosystems CABI Publishing, New York*:27–55.

204. Sanderson MG (1996) Biomass of termites and their emissions of methane and carbon dioxide: A global database. *Global Biogeochem Cycles* 10(4):543–557.

205. Quere CL, et al. (2005) Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Glob Chang Biol* 0(0):051013014052005–???

206. Buitenhuis ET, et al. (2013) MAREDAT: towards a world atlas of MARine Ecosystem DATa. *Earth System Science Data* 5(2):227–239.

207. Buitenhuis ET, Rivkin RB, Sévrine S, Le Quéré C (2010) Biogeochemical fluxes through microzooplankton. *Global Biogeochem Cycles* 24(4). doi:10.1029/2009gb003601.

208. Biard T, et al. (2016) In situ imaging reveals the biomass of giant protists in the global ocean. *Nature* 532(7600):504–507.

209. O'Brien TD (2005) Copepod: A global plankton database. *NOAA Technical Memorandum NMFS-F/SPO-73*:19.

210. Atkinson A, Siegel V, Pakhomov EA, Jessopp MJ, Loeb V (2009) A re-appraisal of the total biomass and annual production of Antarctic krill. *Deep Sea Res Part I* 56(5):727–740.

211. Lucas CH, et al. (2014) Gelatinous zooplankton biomass in the global oceans: geographic variation and environmental drivers. *Glob Ecol Biogeogr* 23(7):701–714.

212. Harris PT, Macmillan-Lawler M, Rupp J, Baker EK (2014) Geomorphology of the oceans. *Mar Geol* 352:4–24.

213. De'ath G, Fabricius KE, Sweatman H, Puotinen M (2012) The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proc Natl Acad Sci U S A* 109(44):17995–17999.

214. Holmes G, Glen H (2008) Estimating three-dimensional surface areas on coral reefs. *J Exp Mar Bio Ecol* 365(1):67–73.

215. Odum HT, Odum EP (1955) Trophic Structure and Productivity of a Windward Coral Reef Community on Eniwetok Atoll. *Ecol Monogr* 25(3):291–320.

216. Vecsei A (2004) A new estimate of global reefal carbonate production including the fore-reefs. *Glob Planet Change* 43(1-2):1–18.

217. McNeil BI (2004) Coral reef calcification and climate change: The effect of ocean warming. *Geophys Res Lett* 31(22). doi:10.1029/2004gl021541.

218. Kuffner IB, Hickey TD, Morrison JM (2013) Calcification rates of the massive coral Siderastrea siderea and crustose coralline algae along the Florida Keys (USA) outer-reef tract. *Coral Reefs* 32(4):987–997.

219. Rodhouse PG, Nigmatullin CM (1996) Role as Consumers. *Philos Trans R Soc Lond B Biol Sci* 351(1343):1003–1022.

220. Buitenhuis ET, et al. (2012) Picophytoplankton biomass distribution in the global ocean. *Earth System Science Data* 4(1):37–46.

221. Leblanc K, et al. (2012) A global diatom database – abundance, biovolume and biomass in the world ocean. *Earth System Science Data* 4(1):149–165.

222. Li WKW, Dickie PM, Irwin BD, Wood AM (1992) Biomass of bacteria, cyanobacteria, prochlorophytes and photosynthetic eukaryotes in the Sargasso Sea. *Deep Sea Res A* 39(3-4):501–519.

223. Agusti S, et al. (2015) Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nat Commun* 6:7608.

224. Herndl GJ (1988) Ecology of amorphous aggregations (marine snow) in the Northern Adriatic Sea. II. Microbial density and activity in marine snow and its implication to overall pelagic processes. *Mar Ecol Prog Ser* 48(3):265–275.

225. Laffoley D, Grimsditch GD (2009) *The Management of Natural Coastal Carbon Sinks* (IUCN).

226. van Elsas JD, Trevors JT, Jansson JK, Nannipieri P (2006) *Modern Soil Microbiology, Second Edition* (CRC Press).

227. Robinson BS, Bamforth SS, Dobson PJ (2002) Density and diversity of protozoa in some arid Australian soils. *J Eukaryot Microbiol* 49(6):449–453.

228. Bamforth SS (2007) Protozoa from aboveground and ground soils of a tropical rain forest in Puerto Rico. *Pedobiologia* 50(6):515–525.

229. Bamforth SS (2008) Protozoa of biological soil crusts of a cool desert in Utah. *J Arid Environ* 72(5):722–729.

230. Berthold A, Palzenberger M (1995) Comparison between direct counts of active soil ciliates (Protozoa) and most probable number estimates obtained by Singh's dilution culture method. *Biol Fertil Soils* 19(4):348–356.

231. Berthold A, Bruckner A, Kampichler C (1999) Improved quantification of active soil microfauna by a "counting crew." *Biol Fertil Soils* 28(4):352–355.

232. Zhao F, Xu K, He Y (2012) Application of the Ludox-QPS method for estimating ciliate diversity in soil and comparison with direct count and DNA fingerprinting. *Eur J Soil Biol* 49(Supplement C):112–118.

233. Roger Anderson O (2015) The role of Bacterial-based Protist Communities in Aquatic and Soil Ecosystems and the Carbon Biogeochemical Cycle, with Emphasis on Naked Amoebae. *Acta Protozool* 51(3):209–221.

234. Adl SM, Coleman DC, Read F (2006) Slow recovery of soil biodiversity in sandy loam soils of Georgia after 25 years of no-tillage management. *Agric Ecosyst Environ* 114(2):323–334.

235. Verhoeven R (2002) The structure of the microtrophic system in a development series of dune soils. *Pedobiologia* 46(1):75–89.

236. Verhoeven R (2002) Ciliates in coastal dune soils of different stages of development. *Eur J Soil Biol* 38(2):187–191.

237. Zhao F, Xu K (2013) Microbial genetic diversity and ciliate community structure along an environmental gradient in coastal soil. *Eur J Protistol* 49(4):516–525.

238. Foissner W (1992) Comparative studies on the soil life in ecofarmed and conventionally farmed fields and grasslands of Austria. *Agric Ecosyst Environ* 40(1):207–218.

239. Wanner M, Xylander WER (2005) Biodiversity development of terrestrial testate amoebae: is there any succession at all? *Biol Fertil Soils* 41(6):428–438.

240. Bobrov AA, Zaitsev AS, Wolters V (2015) Shifts in soil testate amoeba communities associated with forest diversification. *Microb Ecol* 69(4):884–894.

241. Andrey T, Elena E, Yuri M (2012) Distribution of soil testate amoeba assemblages along catenas in the northern taiga zone (Karelia, Russia). *Protistology* 7(2). Available at: https://cyberleninka.ru/article/n/distribution-of-soil-testate-amoeba-assemblages-along-catenas-in-the-northern-taiga-zone-karelia-russia [Accessed December 25, 2017].

242. Nesbitt JE, Adl SM (2014) Differences in soil quality indicators between organic and sustainably managed potato fields in Eastern Canada. *Ecol Indic* 37(Part A):119–130.

243. Carter MR, Gregorich EG (2007) *Soil Sampling and Methods of Analysis, Second Edition* (CRC Press).

244. Stapleton LM, et al. (2005) Microbial carbon dynamics in nitrogen amended Arctic tundra soil: Measurement and model testing. *Soil Biol Biochem* 37(11):2088–2098.

245. Adl SM, Coleman DC (2005) Dynamics of soil protozoa using a direct count method. *Biol Fertil Soils* 42(2):168–171.

246. Finlay BJ, Fenchel T o. m. (2001) Protozoan Community Structure in a Fractal Soil Environment. *Protist* 152(3):203–218.

247. Bamforth SS (1991) Enumeration of soil ciliate active forms and cysts by a direct count method. *Agric Ecosyst Environ* 34(1):209–212.

248. Finlay BJ, et al. (2000) Estimating the growth potential of the soil protozoan community. *Protist* 151(1):69–80.

249. Hengl T, et al. (2014) SoilGrids1km—global soil information based on automated mapping. *PLoS One* 9(8):e105992.

250. Wanner M, Elmer M, Kazda M, Xylander WER (2008) Community assembly of terrestrial testate amoebae: how is the very first beginning characterized? *Microb Ecol* 56(1):43–54.

251. Petz W (1997) Ecology of the active soil microfauna (Protozoa, Metazoa) of Wilkes Land, East Antarctica. *Polar Biol* 18(1):33–44.

252. Allen MF, Klironomos JN, Treseder KK, Oechel WC (2005) RESPONSES OF SOIL BIOTA TO ELEVATED CO2 IN A CHAPARRAL ECOSYSTEM. *Ecol Appl* 15(5):1701–1711.

253. Persson T, et al. (1980) Trophic Structure, Biomass Dynamics and Carbon Metabolism of Soil Organisms in a Scots Pine Forest. *Ecol Bull* (32):419–459.

254. Schaefer M (1990) The soil fauna of a beech forest on limestone: trophic structure and energy budget. *Oecologia* 82(1):128–136.

255. Anderson OR (2008) The role of amoeboid protists and the microbial community in moss-rich terrestrial ecosystems: biogeochemical implications for the carbon budget and carbon cycle, especially at higher latitudes. *J Eukaryot Microbiol* 55(3):145–150.

256. Lousier JD (1982) Colonization of decomposing deciduous leaf litter by Testacea (Protozoa, Rhizopoda): Species succession, abundance, and biomass. *Oecologia* 52(3):381–388.

257. Foissner W (1987) The micro-edaphon in ecofarmed and conventionally farmed dryland cornfields near Vienna (Austria). *Biol Fertil Soils* 3(1-2):45–49.

258. Edgcomb VP, Beaudoin D, Gast R, Biddle JF, Teske A (2011) Marine subsurface eukaryotes: the fungal majority. *Environ Microbiol* 13(1):172–183.

259. Novarino G, et al. (1997) Protistan communities in aquifers: a review. *FEMS Microbiol Rev* 20(3-4):261–275.

260. Griebler C, Lueders T (2009) Microbial biodiversity in groundwater ecosystems. *Freshw Biol* 54(4):649–677.

261. Borgonie G, et al. (2015) Eukaryotic opportunists dominate the deep-subsurface biosphere in South Africa. *Nat Commun* 6:8952.

262. Kinner NE, Harvey RW, Shay DM, Metge DW, Warren A (2002) Field evidence for a protistan role in an organically-contaminated aquifer. *Environ Sci Technol* 36(20):4312–4318.

263. Zhou Y (2013) Spatio-temporal patterns of suspended and attached bacterial communities in a hydrologically dynamic aquifer (Mittenwald, Germany). Dissertation (Technische Universität München). Available at: https://mediatum.ub.tum.de/1120484.

264. Bork P, et al. (2015) Tara Oceans studies plankton at planetary scale. *Science* 348(6237):873–873.

265. Flombaum P, et al. (2013) Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc Natl Acad Sci U S A* 110(24):9824–9829.

266. Antoine D, André J-M, Morel A (1996) Oceanic primary production: 2. Estimation at global scale from satellite (Coastal Zone Color Scanner) chlorophyll. *Global Biogeochem Cycles* 10(1):57–69.

267. Behrenfeld MJ, Falkowski PG (1997) Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol Oceanogr* 42(1):1–20.

268. Wigington CH, et al. (2016) Re-examination of the relationship between marine virus and microbial cell abundances. *Nature Microbiology* 1(3):15024.

269. Engelhardt T, Kallmeyer J, Cypionka H, Engelen B (2014) High virus-to-cell ratios indicate ongoing production of viruses in deep subsurface sediments. *ISME J* 8(7):1503–1509.

270. Kyle JE, Eydal HSC, Grant Ferris F, Pedersen K (2008) Viruses in granitic groundwater from 69 to 450 m depth of the Äspö hard rock laboratory, Sweden. *ISME J* 2(5):571–574.

271. Middelboe M, Glud RN, Filippini M (2011) Viral abundance and activity in the deep sub-seafloor biosphere. *Aquat Microb Ecol* 63(1):1–8.

272. Pan D, Nolan J, Williams KH, Robbins MJ, Weber KA (2017) Abundance and Distribution of Microbial Cells and Viruses in an Alluvial Aquifer. *Front Microbiol* 8:1199.

273. Roudnew B, et al. (2012) Bacterial and virus-like particle abundances in purged and unpurged groundwater depth profiles. *Groundwater Monitoring & Remediation* 32(4):72–77.

274. Williamson KE (2011) Soil Phage Ecology: Abundance, Distribution, and Interactions with Bacterial Hosts. *Biocommunication in Soil Microorganisms*, Soil Biology., ed Witzany G (Springer Berlin Heidelberg), pp 113–136.

275. Parikka KJ, Le Romancer M, Wauters N, Jacquet S (2017) Deciphering the virus-to-prokaryote ratio (VPR): insights into virus-host relationships in a variety of ecosystems. *Biol Rev Camb Philos Soc* 92(2):1081–1100.

276. Shangguan W, Hengl T, Mendes de Jesus J, Yuan H, Dai Y (2017) Mapping the global depth to bedrock for land surface modeling. *J Adv Model Earth Syst* 9(1):65–88.

277. Nigro OD, et al. (2017) Viruses in the Oceanic Basement. *MBio* 8(2). doi:10.1128/mBio.02129-16.

278. Brum JR, et al. (2015) Patterns and ecological drivers of ocean viral communities. *Science* 348(6237):1261498–1261498.

279. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS (2014) The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat Rev Microbiol* 12(7):519–528.

280. Eydal HSC, Jägevall S, Hermansson M, Pedersen K (2009) Bacteriophage lytic to Desulfovibrio aespoeensis isolated from deep groundwater. *ISME J* 3(10):1139–1147.

281. Prangishvili D, et al. (2017) The enigmatic archaeal virosphere. *Nat Rev Microbiol* 15(12):724–739.

282. Sohlenius B (1979) A Carbon Budget for Nematodes, Rotifers and Tardigrades in a Swedish Coniferous Forest Soil. *Holarct Ecol* 2(1):30–40.

283. Milius S, Susan M (2010) Life: First census of marine life reveals there's plenty new under the sea: Decade-long study highlights how much more is left to discover. *Sci News* 178(9):14–14.

284. Gooday AJ, Levin LA, Peter L, Thomas H (1992) The Role of Benthic Foraminifera in Deep-Sea Food Webs and Carbon Cycling. *Deep-Sea Food Chains and the Global Carbon Cycle*, pp 63–91.

285. Rowden AA, et al. (2010) A test of the seamount oasis hypothesis: seamounts support higher epibenthic megafaunal biomass than adjacent slopes. *Mar Ecol* 31:95–106.

286. De Leo FC, Smith CR, Rowden AA, Bowden DA, Clark MR (2010) Submarine canyons: hotspots of benthic biomass and productivity in the deep sea. *Proc Biol Sci* 277(1695):2783–2792.

287. Ichino MC, et al. (2015) The distribution of benthic biomass in hadal trenches: A modelling approach to investigate the effect of vertical and lateral organic matter transport to the seafloor. *Deep Sea Res Part I* 100:21–33.

288. Juniper S, Sarrazin J, Grehan A (1998) Remote sensing of organism density and biomass at hydrothermal vents. *Cah Biol Mar* 39(3-4):245–247.

289. Beaulieu SE, Baker ET, German CR, Andrew M (2013) An authoritative global database for active submarine hydrothermal vent fields. *Geochem Geophys Geosyst* 14(11):4892–4905.

290. Mitarai S, Watanabe H, Nakajima Y, Shchepetkin AF, McWilliams JC (2016) Quantifying dispersal from hydrothermal vent fields in the western Pacific Ocean. *Proc Natl Acad Sci U S A* 113(11):2976–2981.

291. Pella JJ, Tomlinson PK (1969) A generalized stock production model. *Bull IATTC/Bol CIAT* 13(3):416–497.

292. Costello C, et al. (2016) Global fishery prospects under contrasting management regimes. *Proc Natl Acad Sci U S A* 113(18):5125–5129.

293. Thorson JT, Cope JM, Branch TA, Jensen OP (2012) Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. *Can J Fish Aquat Sci* 69(9):1556–1568.

294. Crowther TW, et al. (2015) Mapping tree density at a global scale. *Nature* 525(7568):201–205.

295. Kieft TL, Simmons KA (2015) Allometry of animal–microbe interactions and global census of animal-associated microbes. *Proc R Soc B* 282(1810):20150702.

296. Messager ML, Lehner B, Grill G, Nedeva I, Schmitt O (2016) Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nat Commun* 7:13603.

297. Hatton IA, et al. (2015) The predator-prey power law: Biomass scaling across terrestrial and aquatic biomes. *Science* 349(6252):aac6284.

298. Bakken LR (1985) Separation and purification of bacteria from soil. *Appl Environ Microbiol* 49(6):1482–1487.

299. Veses V, Richards A, Gow NAR (2009) Vacuole inheritance regulates cell size and branching frequency of Candida albicans hyphae. *Mol Microbiol* 71(2):505–519.

300. Bednaršek N, Možina J, Vogt M, O'Brien C, Tarling GA (2012) The global distribution of pteropods and their contribution to carbonate and carbon biomass in the modern ocean. *Earth System Science Data* 4(1):167–186.

301. Fock HO, Ehrich S (2010) Deep-sea pelagic nekton biomass estimates in the North Atlantic: horizontal and vertical resolution of revised data from 1982 and 1983. *J Appl Ichthyol* 26:85–101.

302. Dai L, Li C, Yang G, Sun X (2016) Zooplankton abundance, biovolume and size spectra at western boundary currents in the subtropical North Pacific during winter 2012. *J Mar Syst* 155(Supplement C):73–83.

303. Delia Viñas M, Diovisalvi NR, Cepeda GD (2010) Individual biovolume of some dominant copepod species in coastal waters off Buenos Aires Province, Argentine Sea. *Brazil J Oceanogr* 58(2):177–181.

304. Moriarty R, Buitenhuis ET, Le Quéré C, M.-P. G (2013) Distribution of known macrozooplankton abundance and biomass in the global ocean. *Earth System Science Data* 5(2):241–257.

305. Vogt M, et al. (2012) Global marine plankton functional type biomass distributions: Phaeocystis spp. *Earth System Science Data* 4(1):107–120.

Supplementary Figures and Tables

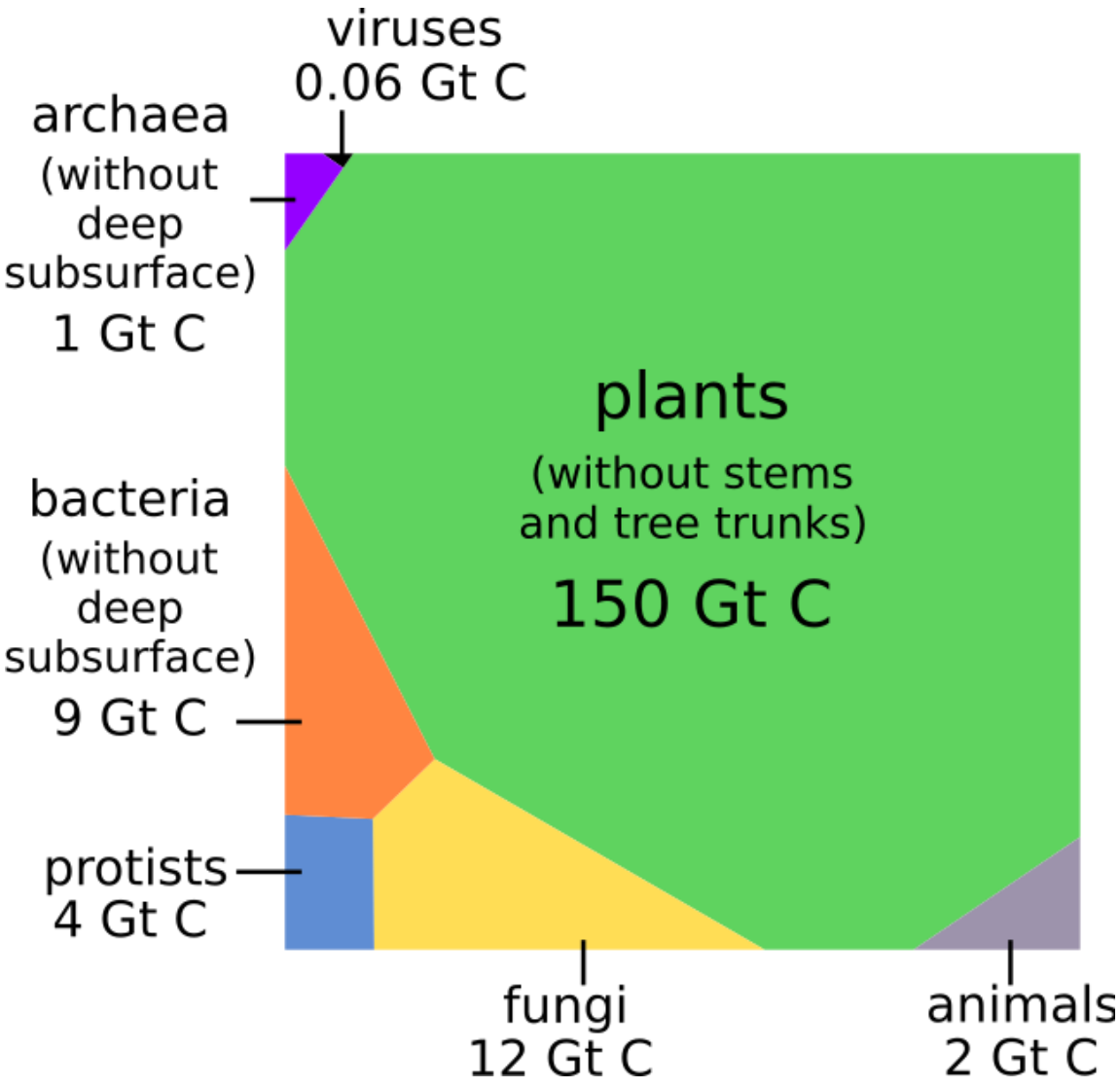


Fig. S1. A graphical representation of the global living biomass distribution.
Absolute living biomass abundances of different taxa are represented in a Voronoi diagram, with the area of each cell being proportional to that taxa global biomass. The specific shape of each cell has no meaning. Values are based on the estimates presented in Table 1. For plants, we exclude the biomass of stems and tree-trunks, as it is mostly composed of non-living lignified tissue. For prokaryotes and viruses, we exclude deep subsurface biomass as the prokaryotes in those environments are metabolically "dormant".

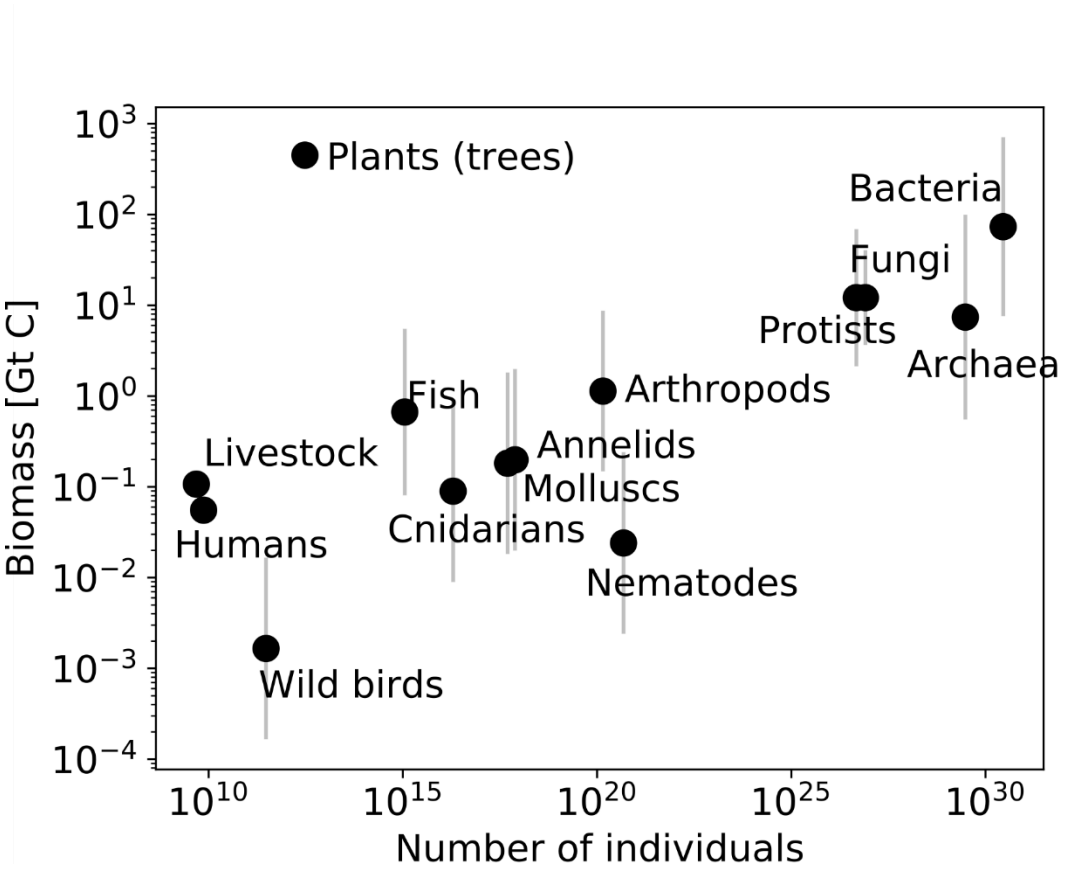


Fig. S2. The relation between abundance and biomass of different taxa.
The total number of individuals in each taxon is plotted against the total biomass of the taxon. The error bars reflect our uncertainty projection of the biomass estimate.

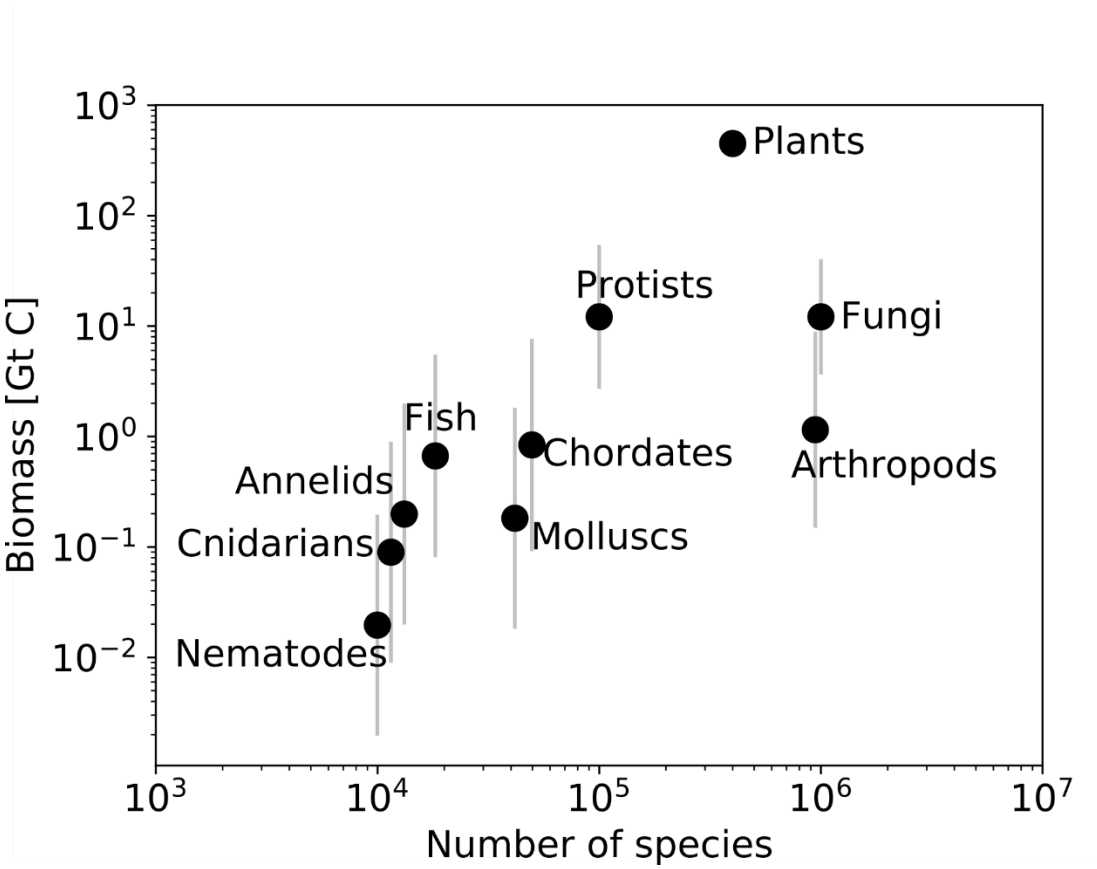


Fig. S3. The relation between species richness and biomass of different taxa.
The total number of species in each taxon is plotted against the total biomass of the taxon. The error bars reflect our uncertainty projection of the biomass estimate. Bacteria, archaea and viruses are not included as the definition of species is problematic. We note that even for fungi and protists the estimate for the number of species may not be robust.

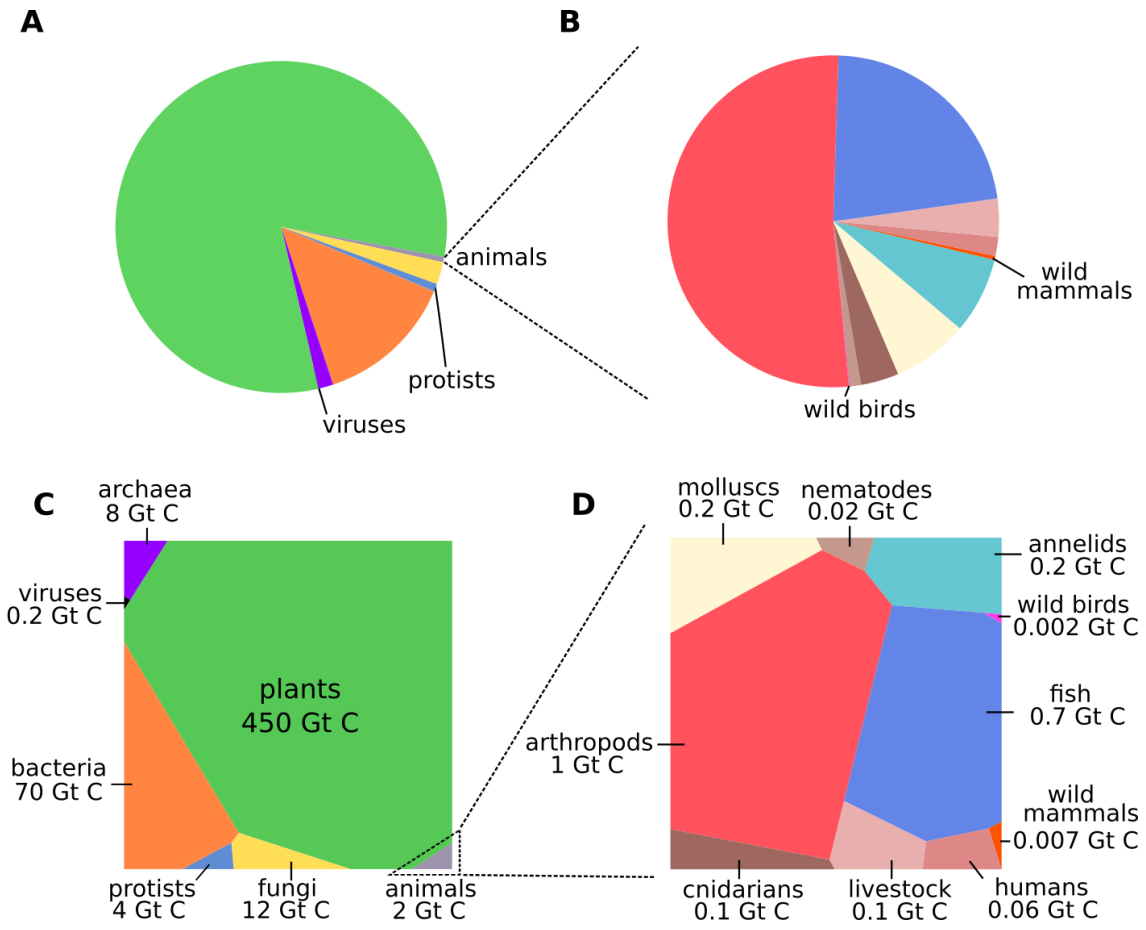


Fig. S4. Comparison of graphical representation of the global biomass distribution using pie-charts versus Voronoi diagrams.

We find that using pie charts that are effectively a one-dimensional presentation modality, the ability to perceive small contributors to the global biomass is severely compromised. (A) Visualization of the distribution of global biomass across the kingdom of life. Compared to the Voronoi diagrams shown in C, due to the inherently lower dynamic range, the contribution of taxa such as protists can hardly be observed and that of viruses is not visible at all. (B) Distribution of biomass focusing on different animal taxa using pie-charts. As observed in the comparison between panels A and C, due to the lower dynamic range of the chart relative to the Voronoi diagram presented in D, the biomass of, for example, wild mammals, is very difficult to discern and that of birds is not visible. Values are based on the estimates presented in Table 1. Only the labels of taxa which are not visible due to the lower dynamic range of the pie-chart are presented in A and B. Colors of each slice correspond to the same taxon as in the Voronoi diagrams.

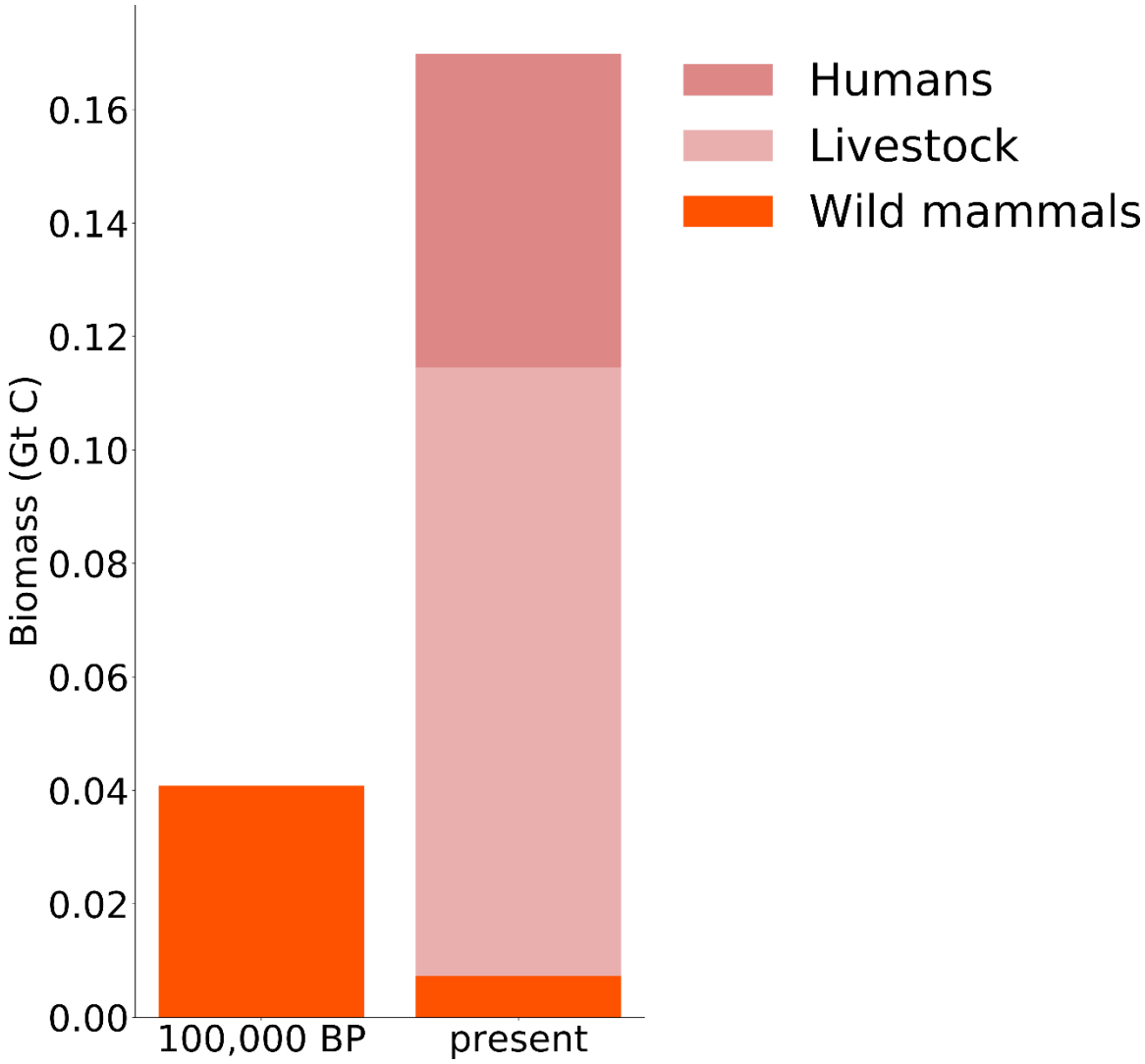


Fig. S5. The impact of human civilization on the biomass of mammals.
The biomass of wild mammals, livestock (dominated by cattle) and humans before human civilization and at present. Values are based on estimates presented in detail in the relevant sections for humans and livestock, wild mammals and pre-human biomass.

Table S1. Summary of estimated total biomass and abundance of various abundant taxonomic groups.

Values are based on a literature survey as detailed in the *SI Appendix*. Reported values have been rounded to the closest order of magnitude in log scale thus reflecting the associated level of uncertainty.

Taxon			Mass [Gt C = 10^{15} g C]		Abundance
Plants	Trees			450	10^{13}
Bacteria	Terrestrial deep subsurface		60		10^{30}
	Marine deep subsurface		7		10^{29}
	Soil		7		10^{29}
	Marine		1.3		10^{29}
	Total			70	10^{30}
Fungi				12	10^{27}
Archaea	Terrestrial deep subsurface		4		10^{29}
	Marine deep subsurface		3		10^{29}
	Soil		0.5		10^{28}
	Marine		0.3		10^{28}
	Total			7	10^{29}
Protists				4	10^{27}
Animals	Chordates	Fish	0.7		10^{15}
		Livestock	0.1		10^{10}
		Humans	0.06		10^{10}
		Wild mammals	0.007		
		Wild birds	0.002		10^{11}
	Arthropods	Terrestrial	0.2		10^{18}
		Marine	1		10^{20}
	Annelids		0.2		10^{18}
	Molluscs		0.2		10^{18}
	Cnidarians		0.1		10^{16}
	Nematodes		0.02		10^{21}
	Total			2	10^{21}
Viruses				0.2	10^{31}

Table S2. Methodology used to estimate the global biomass of plants.
The table summarizes the methods for estimating the biomass of plants. More details are provided in the "Plants – Biomass" section in the *SI Appendix*.

Biomass estimate			
Main methods & resources	Measured quantity	Conversion to biomass	Extrapolation method
Estimate by Erb et al. (9) Based on integration by Erb et al. of both inventory-based data and remote sensing-based data	<u>Inventory based data:</u> Forest area [km^2] and biomass density [g C m^{-2}] Method: data collection in each country worldwide by FRA or based on Pan et al. (1)	Direct biomass measurement	Multiplication of forest area with characteristic biomass densities for each country, done by the FRA or by Pan et al. (1)
	<u>Remote sensing data:</u> Tree height [m] or Growing stock volume [m^3 m^{-2}] Method: Lidar	Conversion made by each study by allometric relations between measured quantity and biomass	Remote sensing measures locally in each location covered. Erb et al. integrated together different studies covering the entire land surface.
Uncertainty			
Variation between different sources which Erb et al. (9) uses for their estimate (link to full calculation)			
Estimate of number of individuals			
We use the estimate of the total number of trees by Crowther et al. (294). We do not give an estimate for the total number of non-tree plants.			

Table S3. Methodology used to estimate the global biomass of marine prokaryotes.
The table summarizes the methods for estimating the biomass of marine bacteria and archaea. More details are provided in the "Bacteria and archaea – marine" section in the *SI Appendix*. The final estimate is a geometric mean of the estimates from Arístegui et al., Buitenhuis et al. and Lloyd et al.

Biomass estimate			
Main Resource	Measured quantity	Conversion to biomass	Extrapolation method
Arístegui et al. (26)	Cell concentration [# cells mL^{-1}] Method: various (meta-analysis), detailed in (26)	Cell carbon content [g C $cell^{-1}$]: 11 fg C, an average of several studies (link to full calculation)	Arístegui et al. calculated average cell concentration for each for depth zones: epipelagic (<200m) mesopelagic (200-1000m) and bathypelagic (1000-4000m). For each zone, Arístegui et al. integrated the biomass density per unit volume over the entire depth of each zone. We calculated the total biomass for each zone by multiplying the above biomass densities by the total surface area of the ocean and summing across all depth zones (link to full calculation)
Buitenhuis et al. (27)	Cell concentration [# cells mL^{-1}] Method: Flow-cytometry and microscopic counting		We binned the data along 100 m depth bins. We calculated the mean concentration of cells at each bin and multiplied the mean cell concentration by the total volume of water at each depth bin
Lloyd et al. (28)	Cell concentration [# cells mL^{-1}] Method: FISH		Lloyd et al. fit an equation predicting cell concentration based on depth. We used the equation to estimate the mean concentration of cells at each depth. We multiplied the mean concentration by the total volume of water at each depth
Uncertainty			
Cell concentration – intra-study and inter-study uncertainty among studies (link to full calculation) Carbon content – intra-study and inter-study uncertainty among several studies (link to full calculation)			
Estimate of number of individuals			
Total number of cells already attained in the process of estimating total biomass			

Table S4. Methodology for estimating the global biomass of soil prokaryotes.
The table summarizes the methods for estimating the biomass of bacteria and archaea in the soil. More details are provided in the "Bacteria and archaea – soil" section in the *SI Appendix*. The final estimate is a geometric mean of the estimates from Xu et al. and Serna-Chavez et al.

<table>
<tr><td colspan="4">Biomass estimate</td></tr>
<tr><td>Main Resource</td><td>Measured quantity</td><td>Conversion to biomass</td><td>Extrapolation method</td></tr>
<tr><td>Serna-Chavez et al. (111)</td><td rowspan="2">Microbial carbon biomass density [g C m^{-2}]
Method: mainly fumigation extraction</td><td rowspan="2">Fraction of prokaryotes out of microbial biomass: ≈0.5 based on Joergensen & Wichern (114)</td><td>Multivariate model by Serna-Chavez et al. (111) predicting biomass density based on mean annual precipitation and temperature, highest annual maximum monthly temperature and soil pH</td></tr>
<tr><td>Xu et al. (110)</td><td>Multivariate model by Xu et al. (110) predicting biomass density based on mean annual precipitation and temperature, and soil organic carbon</td></tr>
<tr><td colspan="4">Uncertainty</td></tr>
<tr><td colspan="4">Microbial biomass – intra-study and inter-study uncertainty across Xu et al. (110) and Serna-Chavez et al. (111; link to full calculation)
Fraction of prokaryotes out of microbial biomass – intra-study, inter-study, inter-habitat (arable, grassland and forest), and inter-method (direct microscopic counts, measurements of specific cell wall components which are characteristic to either fungi or bacteria) uncertainty between all studies reported in Joergensen & Wichern (114; link to full calculation)</td></tr>
<tr><td colspan="4">Estimate of number of individuals</td></tr>
<tr><td colspan="4">Dividing the total biomass of soil prokaryotes by a characteristic cell carbon mass of ≈30 fg C based on Bakken (298)</td></tr>
</table>

Table S5. Methodology for estimating the global biomass of prokaryotes in the marine deep subsurface.
The table summarizes the methods for estimating the biomass of bacteria and archaea in the marine deep subsurface. More details are provided in the "Bacteria and archaea – marine deep subsurface" section in the *SI Appendix*. The final estimate is a geometric mean of the estimates from Kallmeyer et al. and Parkes et al.

<table>
<tr><td colspan="4">Biomass</td></tr>
<tr><td>Main Resources</td><td>Measured quantity</td><td>Conversion to biomass</td><td>Extrapolation method</td></tr>
<tr><td>Kallmeyer et al. (75)</td><td rowspan="2">Cell concentration
[# cells mL^{-1}]
Method: direct counts after DNA staining</td><td rowspan="2">Cell carbon content [g C $cell^{-1}$]: ≈24 fg C, an average of several studies (link to full calculation)</td><td>Multivariate model Kallmeyer et al. (75) predicting cell concentration at each depth based on sedimentation rate and distance from shore. For each depth, Kallmeyer et al. (75) multiplied cell concentration by the total volume at each depth</td></tr>
<tr><td>Parkes et al. (76)</td><td>Parkes et al. (76) used regression of cell concentrations to depth to estimate the concentration of cells at each depth, multiplied cell concentration by the total volume at each depth</td></tr>
<tr><td colspan="4">Uncertainty</td></tr>
<tr><td colspan="4">Number of cells – intra-study and inter-study uncertainty between Kallmeyer et al. and Parkes et al.

Carbon content – intra-study, inter-study and inter-method (volume-based and amino acid-based carbon content estimate) uncertainty across all the studies used to estimate carbon content (link to full calculation)</td></tr>
<tr><td colspan="4">Estimate of number of individuals</td></tr>
<tr><td colspan="4">Total number of cells already attained in the process of estimating total biomass</td></tr>
</table>

Table S6. Methodology for estimating the global biomass of prokaryotes in the terrestrial deep subsurface.

<table>
<tr><td colspan="4">Biomass</td></tr>
<tr><td>Main Resource</td><td>Measured quantity</td><td>Conversion to biomass</td><td>Extrapolation method</td></tr>
<tr><td rowspan="2">McMahon & Parnell (98)</td><td rowspan="2">Cell concentration [# cells mL^{-1}] from groundwater

Method: direct counts after DNA staining</td><td>Cell carbon content [g C cell^{-1}]: ≈26 fg C from McMahon & Parnell (98)</td><td rowspan="2">We use the measurements reported in McMahon & Parnell (98), bin the data to 250 m depth bins, calculate mean cell concentration in each depth bin. We multiply mean cell concentration in each depth by the total volume of aquifer at that depth. Aquifer volume at each depth was estimated by from the data in Gleeson et al. (99) on the total volume of aquifers and porosity profiles</td></tr>
<tr><td>Ratio between number of cells attached to sediment and planktonic cells in groundwater: 10^2-10^3 from McMahon & Parnell (98)</td></tr>
<tr><td colspan="4">Uncertainty</td></tr>
<tr><td colspan="4">Average concentration of cells in groundwater – Based on variance of measurements within each depth bin and difference between the average concentration at each bin based on the geometric mean and the arithmetic mean (link to full calculation)

Aquifer volume – intra-study uncertainty in Gleeson et al. (99)

Carbon content – based on our projection of the uncertainty of the carbon content of bacteria and archaea in subseafloor sediments (see marine deep subsurface bacteria and archaea section)

Attached/Unattached ratio – we use the 95% confidence interval around the geometric mean of the upper and lower bound, 10^2 and 10^3, as our uncertainty range for the ratio of attached to unattached cells.</td></tr>
<tr><td colspan="4">Estimate of number of individuals</td></tr>
<tr><td colspan="4">Total number of cells attained in the process of estimating total biomass</td></tr>
</table>

Table S7. Methodology for estimating the fraction of archaea out of the total biomass of prokaryotes.
The table summarizes the methods for estimating the fraction of archaea out of the total biomass of prokaryotes in each environment. For details, see the relevant section in the supporting materials. Our best estimate is based on a geometric mean of values from independent methods for each environment (Usually FISH and 16S rDNA sequencing).

Environment	Main Resource	Method	Comments
Marine	Lloyd et al. (28)	FISH with specific 16S rDNA probes	We calculated characteristic fraction of archaea in each ocean layer (epipelagic, mesopelagic and bathypelagic)
	Sunagawa et al. (54) Salazar et al. (55)	16S rDNA sequencing	We calculated characteristic fraction of archaea in each ocean layer (epipelagic, mesopelagic and bathypelagic) We corrected for the lower copy number of rDNA genes in archaea (56)
Soil	Bates et al. (58)	16S rDNA sequencing	From 146 soils across the globe. We corrected for the lower copy number of rDNA genes in archaea (56)
	Integration of several studies (link to full calculation)	FISH with specific 16S rDNA probes	We calculated average values across studies and across habitats
	Hong & Cho (59)	16S rDNA qPCR	From grasslands, forests and cropland in Korea
	Integration of several studies (link to full calculation)	CARD-FISH with specific 16S rDNA probes	We calculated average values across studies and across habitats
Marine deep subsurface	Lloyd et al. (28)	CARD-FISH with specific 16S rDNA probes	We calculated the average over all samples taken in layers deeper than 10 cm
Terrestrial deep subsurface	Several studies (101–104) (link to full calculation)	16S rDNA sequencing	We corrected for the lower copy number of rDNA genes in archaea (56) We calculated average values across studies
	Several studies (105–107) (link to full calculation)	16S rDNA qPCR	We calculated average values across studies
	Moser et al. (108)	FISH with specific 16S rDNA probes	
	Lloyd et al. (28)	CARD-FISH with specific 16S rDNA probes	We used our best estimate for the fraction of archaea in the marine deep subsurface as an additional source for estimating the fraction of archaea in the terrestrial deep subsurface

Table S8. Methodology for estimating the global biomass of fungi.
The table summarizes the methods for estimating the biomass of fungi. More details are provided in the "Fungi" section in the supporting materials. The final estimate is a geometric mean of the estimates from Xu et al. and Serna-Chavez et al.

<table>
<tr><td colspan="4">Biomass estimate</td></tr>
<tr><td>Main Resource</td><td>Measured quantity</td><td>Conversion to biomass</td><td>Extrapolation method</td></tr>
<tr><td>Serna-Chavez et al. (111)</td><td rowspan="2">Microbial carbon biomass density [g C m^{-2}]
Method: mainly fumigation extraction</td><td rowspan="2">Fraction of prokaryotes out of microbial biomass: ≈0.5 based on Joergensen & Wichern (114)</td><td>Multivariate model by Serna-Chavez et al. (111) predicting biomass density based on mean annual precipitation and temperature, highest annual maximum monthly temperature and soil pH</td></tr>
<tr><td>Xu et al. (110)</td><td>Multivariate model by Xu et al. (110) predicting biomass density based on mean annual precipitation and temperature, and soil organic carbon</td></tr>
<tr><td colspan="4">Uncertainty</td></tr>
<tr><td colspan="4">Microbial biomass – intra-study and inter-study uncertainty across Xu et al. (110) and Serna-Chavez et al. (111; link to full calculation)
Fraction of prokaryotes out of microbial biomass – intra-study, inter-study, inter-habitat (arable, grassland and forest), and inter-method (direct microscopic counts, measurements of specific cell wall components which are characteristic to either fungi or bacteria) uncertainty between all studies reported in Joergensen & Wichern (114; link to full calculation)</td></tr>
<tr><td colspan="4">Estimate of number of individuals</td></tr>
<tr><td colspan="4">Dividing the total biomass of soil prokaryotes by a characteristic cell carbon mass of ≈15 pg C based on a rough estimate of a cell volume of ≈100 μm^3 (299) and using a carbon density of ≈150 fg C μm^{-3}</td></tr>
</table>

Table S9. Methodology for estimating the global biomass of humans.
The table summarizes the methods for estimating the biomass of Humans. More details are provided in the "Humans and livestock" section in the supporting materials.

Biomass estimate		
Main Resource	Measured quantity	Conversion to biomass
UN World Population Prospects, 2015 revision	Total number of humans	Characteristic carbon content per person: 7500 g C based on Hern (172) assuming 70% water content and 50% carbon content of dry weight (age weighted average body weight is ≈50 kg)
Estimate of number of individuals		
Total number of individuals attained in the process of estimating total biomass		

Table S10. Methodology for estimating the global biomass of livestock.
The table summarizes the methods for estimating the biomass of livestock. More details are provided in the "Humans and livestock" section in the supporting materials.

Biomass estimate			
Main Resource	Measured quantity	Conversion to biomass	Extrapolation method
FAOStat database (domain: Production/Live animals)	Total number of livestock	Characteristic weight of each livestock species in each continent based on IPCC (171). We convert weight to carbon assuming 70% water content and 50% carbon mass out of dry weight	We sum of total number of individuals in each country, and multiply it by the characteristic carbon content of each species of livestock
Estimate of number of individuals			
Total number of individuals attained in the process of estimating total biomass			

Table S11. Methodology for estimating the global biomass of wild mammals.
The table summarizes the methods for estimating the biomass of wild mammals. More details are provided in the "Humans and livestock" section in the supporting materials. Our best estimate for the biomass of wild mammals is the sum of our estimates for terrestrial and marine wild mammals.

Biomass estimate				
Main Resources		Measured quantity	Conversion to biomass	Extrapolation method
Wild land mammals (geometric mean of the thee approaches)	Smil (173)	Characteristic dry mass densities per biome [kg ha^{-1}] from the HYDE database (174)	Assuming 70% water content and 50% carbon content out of dry weight	Interpolation of characteristic mass densities across the area of each biome, done by Smil (173)
	Novosolov et al. (176), IUCN data	Mass of individual [g] and range [m^2] for each species		Log-log correlations between the total number of individuals, the mass of an individual, and the study area in which the total number of individuals were measured were established. Based on these correlations, the total number of individuals was extrapolated based on range size data and body mass data. The extrapolated total number of individuals was multiplied by the mass of an individual to generate the estimate of total biomass.
	Barnosky (175)	Number density of animals and ranges, done by Barnosky (175)		Based on relation between animal mass and its population density, done by Barnosky (175)
Marine mammals	Christensen (178)	Estimates from Christensen (178) of the total number of individuals combined with estimates of the characteristic weights of individuals in Christensen (178). We compared the estimates against data from the IUCN (link to full calculation)		

Table S12. Methodology for estimating the global biomass of wild birds.
The table summarizes the methods for estimating the biomass of birds. More details are provided in the "Birds" section in the supporting materials. Our best estimate for the biomass of birds is the geometric mean of the estimates using methods 1 & 2.

Biomass estimate				
Main methods & resource		Measured quantity	Conversion to biomass	Extrapolation method
Method 1	Gaston & Blackburn (180)	Density of number of birds [# individuals m^{-2}] Methods: several, detailed in Gaston & Blackburn (180)	We calculated a characteristic carbon content of individual bird based on Nee et al. (181). We used this carbon content to convert the total number of birds to total biomass	Gaston & Blackburn extrapolated the density of birds per unit area to the total surface area of ice-free land.
Method 2	Novosolov et al. (176)	Mass of individual [g] (177) and range [m^2] for each species (176)	Assuming 70% water content and 50% carbon content out of dry weight	Log-log correlations between the total number of individuals, the mass of an individual, and the study area in which the total number of individuals were measured were established. Based on these correlations, the total number of individuals was extrapolated based on range size data and body mass data. The extrapolated total number of individuals was multiplied by the mass of an individual to generate the estimate of total biomass.
Estimate of number of individuals				
Total number of individuals attained in the process of estimating total biomass				

Table S13. Methodology for estimating the global biomass of marine arthropods.
The table summarizes the methods for estimating the biomass of marine arthropods. More details are provided in the "Marine arthropods" section in the supporting materials. Our best estimate for the biomass of marine arthropods is the sum of the biomass of arthropods in the mesozooplakton and macrozooplankton. For mesoplankton and microzooplankton, our estimates based on the geometric mean between estimates are based on the average and median biomass densities.

Biomass estimate				
Main Resources		Measured quantity	Conversion to biomass	Extrapolation method
mesozooplankton	Buitenhuis et al. (206) - assuming crustaceans dominate mesozooplankton	Direct biomass density or number of individuals density [g C mL^{-1}]/ [# of individuals mL^{-1}] Method: various (meta-analysis)	If biomass not measured directly, Buitenhuis et al. used length of individuals and allometric equations between length and biomass	Buitenhuis et al. calculated average/median concentration per depth integral and multiplied it by the total volume of each depth layer
macrozooplankton	Moriarty et al. (304)	Direct biomass density or number of individual density [g C mL^{-1}]/ [# of individuals mL^{-1}] Method: various (meta-analysis)	If biomass not measured directly, Moriarty et al. used length of individuals and allometric equations between length and biomass	Moriarty et al. calculated average/median concentration per depth integral and multiplied it by the total volume of each depth layer. We subtract the contribution from other dominant macrozooplankton taxa such as cnidarians, molluscs etc.
Uncertainty				
As our estimate relies on the MAREDAT database, we use our projection for the uncertainty of the database, which is based on a dedicated section above titled "Sanity checks on the MAREDAT dataset				
Estimate of number of individuals				
Based on dividing the total biomass of each size category (mesozooplankton/macrozooplankton) by the characteristic carbon content of each class. The carbon content is calculated by taking the geometric mean of the values from the literature (382, 303)				

Table S14. Methodology for estimating the global biomass of terrestrial arthropods.
The table summarizes the methods for estimating the biomass of terrestrial arthropods. More details are provided in the "Terrestrial arthropods" section in the supporting materials. Our best estimate for the biomass of terrestrial arthropods is the geometric mean of the two methods.

<table>
<tr><td colspan="4">Biomass estimate</td></tr>
<tr><td colspan="2">Main Methods & Resources</td><td>Measured quantity</td><td>Conversion to biomass</td><td>Extrapolation method</td></tr>
<tr><td>Method 1</td><td>Various sources (link to full calculation)</td><td>Density of number of individuals [# individuals m^{-2}]

Method: direct counts and length measurements</td><td>Length of individuals and allometric equations between length and biomass</td><td rowspan="2">Assuming the biomass density is applicable world-wide. We multiply the average biomass density by the total non-ice surface of the world</td></tr>
<tr><td>Methods 2</td><td>Williams (195)</td><td>Density of number of individuals [# individuals m^{-2}]

Method: direct counts, measured in southeast England</td><td>Characteristic mass calculated from method 1 by dividing the total biomass in each site by the total number of individuals reported</td></tr>
<tr><td colspan="4">Uncertainty</td></tr>
<tr><td colspan="4">Average insect biomass – intra-study and inter-study uncertainty between different studies

Total number of individuals - the uncertainty reported in the paper by Williams (195) for the range of total number of insects</td></tr>
<tr><td colspan="4">Estimate of number of individuals</td></tr>
<tr><td colspan="4">We use the total number of individuals estimated by Williams (195)</td></tr>
</table>

Table S15. Methodology for estimating the global biomass of fish.
The table summarizes the methods for estimating the biomass of fish. More details are provided in the "Fish" section in the supporting materials. Our best estimate for the biomass of fish integrates our estimate for mesopelagic fish and non-mesopelagic fish.

<table>
<tr><td colspan="6">Biomass estimate</td></tr>
<tr><td colspan="3">Main Methods & Resources</td><td>Measured quantity</td><td>Conversion to biomass</td><td>Extrapolation method</td></tr>
<tr><td rowspan="3">Mesopelagic fish (geometric mean of both methods)</td><td rowspan="2">Acoustics based (geometric mean of both studies)</td><td>Irigoien et al. (164)</td><td rowspan="2">area backscattering coefficient [m^2 m^{-2}]
Method: Simrad EK60 echosounder operating at 38 kHz</td><td rowspan="2">We use target strength per biomass measurements from Irigoien et al. for fish with and without swimbladder and assume ≈50% of the population has swimbladder</td><td>Irigoien et al. used correlation between the scattering coefficient and primary productivity</td></tr>
<tr><td>Proud et al. (165)</td><td>Proud et al. calculated mean backscatter for ocean areas clustered based on primary productivity, and multiplied the mean backscatter by the area of each ocean area</td></tr>
<tr><td>Trawling based</td><td>Lam & Pauly (163)</td><td>Wet weight [g m^{-2}] density

Method: trawling</td><td>Assuming water content of ≈70% and carbon content of ≈50% of dry weight</td><td>Lam & Pauly multiply average densities with the surface areas for different strata of the world ocean</td></tr>
<tr><td colspan="2">non-mesopelagic fish</td><td>Wilson et al. (161)</td><td colspan="3">Ecological modelling by Wilson et al. based on various parameters such as primary productivity, transfer efficiencies etc.</td></tr>
<tr><td colspan="6">Uncertainty</td></tr>
<tr><td colspan="2">Mesopelagic fish</td><td colspan="4">Total scattering – intra-study and inter-study uncertainty between Irigoien et al. and Proud et al.

Target strength per biomass - the main parameter affecting the target strength is the presence of absence of a swimbladder. Testing the sensitivity of the estimate to different population compositions of fish containing or lacking swimbladder.

The total fraction of mesopelagic scattering layer that is fish - assumed to be high based on Kloser et al. and Godø et al. (168, 169)

Other sources of uncertainty - the influence of the size spectra of fish on the target strength of a single fish, possible effects of resonance</td></tr>
<tr><td colspan="2">Non-mesopelagic fish</td><td colspan="4">Intra-study uncertainty reported in a similar ecological model by Jennings et al. (170)</td></tr>
<tr><td colspan="6">Estimate of number of individuals</td></tr>
<tr><td colspan="6">Dividing the total biomass of mesopelagic fish by a characteristic carbon content of a single fish of ≈0.5 g C, based on Fock & Ehrich (301)</td></tr>
</table>

Table S16. Methodology for estimating the global biomass of annelids.
The table summarizes the methods for estimating the biomass of annelids. More details are provided in the "Annelids" section in the supporting materials. Our estimate is based on the geometric mean between estimates based on the average and median biomass densities.

Biomass estimate			
Main Resource	Measured quantity	Conversion to biomass	Extrapolation method
Fierer et al. (144)	Abundance density [# on individuals m^{-2}] Method: direct counts	Characteristic carbon content of earthworms: 5 mg C based on Petersen & Luxton (146)	We multiplied average densities per biome from Fierer et al. (144) by area of each biome
Estimate of number of individuals			
Total number of individuals attained in the process of estimating total biomass			

Table S17. Methodology for estimating the global biomass of molluscs.
The table summarizes the methods for estimating the biomass of molluscs. More details are provided in the "Molluscs" section in the supporting materials. Our best estimate for the biomass of molluscs is the sum of the estimates for pteropods and squids. For pteropods, our estimates are based on the geometric mean between estimates based on the average and median biomass densities.

Biomass estimate				
Main Resource		Measured quantity	Conversion to biomass	Extrapolation method
Pteropods	Bednaršek et al. (300)	Either biomass density or individual abundance density [g C m^{-3}]/[# individuals m^{-3}] Methods: various (meta-analysis)	In case biomass was not directly measured, length of the organism was used to convert to biomass using allometric equations, as detailed in Bednaršek et al. (300)	Bednaršek et al calculated average/median concentration per depth integral and extrapolating based on the total volume of each depth layer
Squids	Rodhouse & Nigmatullin (219)			
Estimate of number of individuals				
We rely on estimates of the average abundance and biomass densities of pteropods (300). We divide the two to get a characteristic biomass per individual. We then divide the total biomass of pteropods by the characteristic biomass of an individual. Pteropods are small compared to squids, so they will dominate the abundance of molluscs.				

Table S18. Methodology for estimating the global biomass of cnidarians.
The table summarizes the methods for estimating the biomass of cnidarians. More details are provided in the "Cnidarians" section in the supporting materials. Our best estimate for the biomass of cnidarians is the sum of the estimates for cnidarians in plankton and for corals.

<table>
<tr><td colspan="5">Biomass estimate</td></tr>
<tr><td colspan="2">Main methods & resource</td><td>Measured quantity</td><td>Conversion to biomass</td><td>Extrapolation method</td></tr>
<tr><td>Gelatinous plankton</td><td>Lucas et al. (211)</td><td>Abundance data for each taxon [# individuals m^{-3}]

Method: various (meta-analysis), detailed in Lucas et al. (211)</td><td>Lucas et al. used characteristic length measurements for each taxon and allometric equations from length to biomass from various sources listed in (211)</td><td>Lucas et al. applied mean concentration of cnidaria across the top 200 meters of the ocean</td></tr>
<tr><td rowspan="4">Corals

(Geometric mean of method 1 & 2)</td><td rowspan="3">Method 1</td><td rowspan="3">Tissue volume density [mL m^{-2}]

Method: vaseline method based on Odum & Odum (215)</td><td rowspan="3">Assuming carbon density similar to that of sea anemones based on Odum & Odum (215)</td><td>We assume coral tissue surface area is ≈5 times the projected area based on Holmes & Glen (214)</td></tr>
<tr><td>We assume corals occupy ≈20% of coral reef area based on De'ath et al. (213)</td></tr>
<tr><td>We multiply biomass density by the area of coral reefs based on Harris et al. (212) - accounting for coral cover and tissue surface area</td></tr>
<tr><td rowspan="2">Method 2</td><td rowspan="2">Local carbonate production [kg carbonate m^{-2} $year^{-1}$]

Methods: various based on McNiel (217) and Kuffner et al. (218)</td><td>Carbonate production per unit surface area of coral [kg carbonate $year^{-1}$ m^{-2}] Based on McNiel (217) and Kuffner et al. (218)</td><td rowspan="2">Applying local production per area to the total area of coral reefs based on Vecsei (216)</td></tr>
<tr><td>Carbon content per unit surface area from method 1</td></tr>
<tr><td colspan="5">Estimate of number of individuals</td></tr>
<tr><td colspan="5">Dividing total biomass estimate for gelatinous plankton by a characteristic carbon content per individual based on Lucas et al. (211)</td></tr>
</table>

Table S19. Methodology for estimating the global biomass of nematodes.
The table summarizes the methods for estimating the biomass of nematodes. More details are provided in the "Nematodes" section in the supporting materials. Our estimate is based on the geometric mean between estimates based on the average and median biomass densities.

Biomass estimate			
Main Resource	Measured quantity	Conversion to biomass	Extrapolation method
Fierer et al. (144)	Abundance density [# on individuals m^{-2}] Method: direct counts	Characteristic carbon content of nematode: ≈0.05 µg C based on Petersen & Luxton (146)	We multiplied average densities per biome from Fierer et al. (144) by area of each biome
Estimate of number of individuals			
Total number of individuals attained in the process of estimating total biomass			

Table S20. Methodology for estimating the global biomass of marine protists.
The table summarizes the methods for estimating the biomass of marine protists. More details are provided in the "Marine protists" section in the supporting materials. Our best estimate for the biomass of marine protists is the sum of the estimates for each size fraction. For picophytoplankton, and microzooplankton, diatoms, *Phaeocystis* and macrozooplankton, our estimates are based on the geometric mean between estimates based on the average and median biomass densities.

Biomass estimate				
Main size fractions & resource		Measured quantity	Conversion to biomass	Extrapolation method
Picophytoplankton	Buitenhuis et al. (206, 220)	Density of number of cells [# cells m^{-3}] Method: various (meta-analysis), detailed in (206)	Buitenhuis et al. used characteristic carbon content for an individual cell from various sources listed in (206, 220)	Buitenhuis et al. calculated average/median concentration per depth integral and multiplied it by the total volume of each depth layer
Heterotrophic Pico-nanoplankton	de Vargas et al. (130)	Fraction of pico-nanoplankton which are autotrophic Method: 18S metabarcoding	We take our estimate for picophytoplankton and divide it by the the ratio of autotrophic to heterotrophic pico-nanoplankton	
microzooplankton	Buitenhuis et al. (206)	Direct biomass density or number of individual density [g C mL^{-1}]/ [# of individuals mL^{-1}] Method: various (meta-analysis), detailed in (206)	If biomass not measured directly, Buitenhuis et al. used length of individuals and allometric equations between length and biomass	Buitenhuis et al. calculated average/median concentration per depth integral and multiplied it by the total volume of each depth layer
Diatoms	Leblanc et al. (221)	Density of number of cells [# cells m^{-3}] Method: various (meta-analysis), detailed in (221)	Leblanc et al. used characteristic carbon content for an individual cell from various sources listed in (221)	Leblanc et al. calculated average/median concentration per depth integral and multiplied it by the total volume of each depth layer

Phaeocystis	Vogt et al. (307)	Density of number of cells [# cells m^{-3}] Method: various, mostly microscopy	Vogt et al. used characteristic carbon content for an individual cell from various sources listed in (307)	Vogt et al. calculated average/median concentration per depth integral and multiplied it by the total volume of each depth layer
Macrozooplankton	Biard et al. (208)	Biovolume density [m^{-3} m^{-3}] Method: *in situ* imaging	Biard et al. used characteristic carbon content for an individual cell from various sources listed in Biard et al. (208)	Biard et al. calculated average/median concentration per depth integral and multiplied it by the total volume of each depth layer
Uncertainty				
As our estimate relies on the MAREDAT database, we use our projection for the uncertainty of the database, which is based on a dedicated section above titled “Sanity checks on the MAREDAT dataset				
Estimate of number of individuals				
Divide total biomass of each size category by the characteristic carbon content of each class. The carbon content is calculated by taking the geometric mean of the volume of each size category, and using a conversion equation between biovolume and carbon content from Pernice et al. (132)				

Table S21. Methodology for estimating the global biomass of terrestrial protists.
The table summarizes the methods for estimating the biomass of terrestrial protists. More details are provided in the "Terrestrial protists" section in the supporting materials.

Biomass estimate			
Main Resource	Measured quantity	Conversion to biomass	Extrapolation method
Integration of several studies (link to full calculation)	Density of number of individuals [# individuals m^{-3}] Method: direct microscopical counts	We use an average carbon content for each morphological type (flagellates, ciliates, and naked and testate amoebae) from values reported in several studies (link to full calculation)	We calculate mean individual density for each biome, multiply the mean density by the total area of each biome and sum across all biomes.
Uncertainty			
Number of individuals – intra-study and inter-study uncertainty between the different studies. Inter-habitat uncertainty between mean density of individuals in different habitats (link to full calculation) Carbon content – intra-study and inter-study uncertainty between different studies amoebae (link to full calculation)			
Estimate of number of individuals			
Total number of individuals attained in the process of estimating total biomass			

Table S22. Methodology for estimating the global biomass of viruses.
The table summarizes the methods for estimating the biomass of viruses. More details are provided in the "Viruses" section in the supporting materials. Our best estimate for the biomass of viruses integrates our estimate for the biomass of viruses in each environment.

Biomass estimate				
Main environments & resource		Measured quantity	Conversion to biomass	Extrapolation method
Marine	Wigington et al. (268)	Population density of phages [# virions mL^{-1}] Methods: direct counts	Assuming a characteristic size of phages from Brum et al. (278) and a carbon content per phage size from Jover et al. (279)	We bin the data to different depth bins, calculate the mean density of individuals in each bin and multiply the mean density by the total volume of water in the bin
Soil	Williamson (274) Parikka et al. (275)	Population density of phages [# individuals g^{-1}] Methods: various (meta-analysis), detailed in (274, 275)		We use an average phage density from Williamson (274) and Parikka et al. (275) and extrapolate it across the entire volume of soil
Marine subsurface	Engelhardt et al. (269)	Phage to prokaryotes ratio Methods: direct counts		We use our estimate for the total number of marine prokaryotes and multiply it by the geometric mean of the phage to prokaryotes ratio from Engelhardt et al.
Terrestrial subsurface	Engelhardt et al. (269), Kyle et al. (270), Pan et al. (272), Roundnew et al. (273)	Phage to prokaryotes ratio Methods: direct counts		We use our estimate for the total number of marine prokaryotes and multiply it by the phage to prokaryotes ratio. We generate several estimates based on different assumptions and take the geometric mean of all of them. For more details see the relevant sections above
Uncertainty				
Number of viruses – For each environment we calculate intra-study and interstudy uncertainty. We propagate the uncertainties in each environment to our total sum of the number of virions Size of phages – Intra-study uncertainty from Brum et al. (278). In line with measurements of phage size from terrestrial deep subsurface (280) Carbon content of phages – Uncertainty of estimate of the biophysical model in Jover et al. (279)				
Estimate of number of individuals				
Total number of individuals attained in the process of estimating total biomass				

Table S23. The global biomass concentrated in the terrestrial, marine or deep subsurface environments.

The table details the taxa and considered for calculating the global biomass in the terrestrial, marine, and deep subsurface environments.

Terrestrial		Marine		Deep subsurface	
Taxon	Biomass [Gt C]	Taxon	Biomass [Gt C]	Taxon	Biomass [Gt C]
Plants	450	Marine bacteria	1.3	Terrestrial deep subsurface bacteria	60
Soil fungi	12	Marine protists	2	Marine deep subsurface bacteria	7
Soil bacteria	7	Marine arthropods	1	Terrestrial deep subsurface archaea	4
Terrestrial protists	1.6	Fish	0.7	Marine deep subsurface archaea	3
Soil archaea	0.5	Marine fungi	0.3		
Terrestrial arthropods	0.2	Marine archaea	0.3		
Annelids	0.2	Marine molluscs	0.2		
Livestock	0.1	Cnidaria	0.1		
Humans	0.06	Marine nematodes	0.01		
Wild mammals	0.007				
Terrestrial nematodes	0.006				
Wild birds	0.002				
Sum	470	Sum	6	Sum	70