The biomass composition of the oceans - a blueprint of our blue planet

Yinon M. Bar-On & Ron Milo

Supplementary information

Table of Contents

Here we provide a detailed description of the data and procedures we used to arrive at the final estimates presented in the paper. The description is divided into different groups of organisms. For some organisms, there is a further division into the different environments they reside in. We used **bold font** to highlight the final values derived for each taxonomic group. All 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/ocean_biomass.

The supplementary information 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 will be able to update the analysis using extra data or a different data analysis approach. To generate our estimates of biomass, we extracted values from the literature into spreadsheet files. Our analysis pipeline is comprised of about 6 different Jupyter notebooks, which use the data extracted from the literature as input and generate our estimates. The results of our analysis are summarized in a summary table located at the following link. We use the results of our analysis when reporting the values in the manuscript, including all the associated figures.

As a general rule, the data for this analysis is based on the integration of sources performed in (Bar-On et al., 2018). For further details on the specific procedures for estimating the biomass of each group, as well as the uncertainty associated with each estimate, see the SI appendix of (Bar-On et al., 2018). We list below specific instances in which we updated values from (Bar-On et al., 2018) based on recent data.

Usage of various estimators of the mean

We followed a similar approach to that used in (Bar-On et al., 2018). For the sake of the coherence of the supplementary information, we reiterate our methodology below. In order to generate global estimates based on local samplings, we calculate the average values from all the peer-reviewed literature sources we could find and extrapolate from them to the global scale. We calculated two types of averages. The first is the characteristic value from different independent measurements, for example, calculating the characteristic carbon content of bacteria from different studies that 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. The variability in reported values across different studies is often the result of different measurement methodologies or biases rather than actual differences in the populations sampled. In such cases, we used the geometric mean of the values from the different studies as our best estimate, because we assumed, for lack of better knowledge, that the measurement error, in this specific example, is the measurement error of the carbon content of bacteria, which tends to be 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 calculated is when we used different samples of the population density of organisms in the environment and calculated 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 disadvantages, we chose throughout our analysis to calculate the average population density based on both the arithmetic and 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 used 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 more standard approach could be to use the median over the reported values, but this method will disregard the effect of locations with especially high population density. We note that the statistical approach we chose to use is ad-hoc and not a formal approach based on a rigorous underlying statistical process model. We selected it as its relative simplicity makes the workflow more transparent than more rigorous approaches, which are much harder to follow.

Uncertainty analysis

We followed a similar approach for estimating uncertainty as that used in (Bar-On et al., 2018). For the sake of the coherence of the supplementary information, we describe the method from (Bar-On et al., 2018). Alongside describing the procedures leading to the estimates of the parameters used to derive the biomass composition in the oceans, we quantitatively survey the main sources of uncertainty associated with each parameter and calculate an uncertainty range for each. We chose to report uncertainties as representing, to the best of our ability given the many constraints, biases and limitations, the equivalent of the 95% confidence interval for the estimate of the mean. Uncertainties reported in our analysis are multiplicative (fold change from the mean) and not additive (± change of the estimate). We chose to use multiplicative uncertainty because it is more robust to possible outliers in the underlying data and because it is a natural way to report uncertainties associated with the geometric mean of a sample. We calculated the uncertainty of each quantity around the geometric mean of the data used to estimate it by taking the logarithm of the values reported either within studies or from different studies. Taking the logarithm moves the values to log-space, where the SE is calculated (by dividing the SD by the square root of the number of values). We then multiplied the SE by a factor of 1.96, which will give the 95% confidence interval if the transformed data are normally distributed. Finally, we exponentiated the result to get the multiplicative factor in linear space that represents the confidence interval (akin to a 95% confidence interval if the data are lognormally distributed). When data is ample, the uncertainty around the geometric mean will be low (as we base our uncertainty on the SE). Nevertheless, this type of uncertainty does not consider the possibility that the distribution of values in the sample data does not represent the natural environment faithfully, which is probably common in measurements that are sparse and biased.

Many of our estimates are constructed by combining several different estimates (for example, combining the total number of individuals with the characteristic carbon content of a single organism). In these cases, we used intra-study, inter-study or inter-method variations associated with each of the parameters used to derive the final estimate, and propagated these uncertainties to the final estimate of biomass. The

uncertainty analysis for each specific biomass estimate incorporates different components of this general scheme, depending on the amount of information available for estimating the uncertainty of the specific biomass, as we discuss on a case-by-case basis below.

In cases where information is ample, the procedure described above yields several different uncertainty estimates for each parameter used to derive the final estimate (for example, intra-study and inter-study uncertainty). We integrate these different uncertainties together, usually by taking the highest value among them as our best projection of the uncertainty associated with the estimated parameter. In cases in which information was scarce or some sources of uncertainty were hard to quantify, we also used our subjective judgment based on reading the available literature and consulting with experts in each field of study, when available. We tend to be conservative and round up our uncertainty projections when data is especially limited.

Phytoplankton

In this section, we estimate the global mass of phytoplankton (autotrophic plankton) in the ocean, as well as the partial contribution of bacterial phytoplankton (cyanobacteria), protist phytoplankton, and green microalgae to the global biomass of phytoplankton. We relied on two general approaches for estimating the global mass of phytoplankton. The first is a bottom-up approach, which estimates the global biomass of the major components of phytoplankton separately. The second is a top-down approach, in which we estimated the global biomass of phytoplankton collectively and then divided it into its main constituents. The total biomass of each taxonomic group of phytoplankton (bacterial, plant and protist phytoplankton) is then considered as part of the total biomass of each kingdom in their respective sections.

Bottom-up approach

Our bottom-up approach is based on the MAREDAT initiative, which includes estimates for picophytoplankton (Buitenhuis et al., 2012a, 2013), diatoms (Leblanc et al., 2012), and *Phaeocystis* (Buitenhuis et al., 2013). 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 it 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.

Estimates for the global picophytoplankton biomass (Buitenhuis et al., 2012a, 2013) converge at ≈0.4 Gt C. Buitenhuis et al. (Buitenhuis et al., 2012a, 2013) estimates that picoeukaryotes represent ≈60% (49-69%) of the global biomass of picophytoplankton, which translates to ≈0.25 Gt C of picoeukaryotes (link to full calculation), with the remaining 0.17 Gt C contributed by cyanobacteria. Picoeukaryotes contain both protists and plant species (like chlorophytes). A recent study suggests green algae account for ≈20% of the biomass of picoeukaryotes (Limardo et al., 2017). A major caveat of this study is that it looks only at two dominant species of chlorophytes, namely, *Ostreococcus* and *Bathycoccus*, and thus might underestimate the contribution of chlorophytes to picophytoplankton. On the other hand, this study only looks at the abundance of cells and not at the biomass contribution of chlorophytes. As *Bathycoccus* and, more specifically, *Ostreococcus* are known to be small, looking only at their population size might overestimate their biomass contribution. Proceeding in spite of this major caveat, we estimate the global biomass of protist picoeukaryotes at ≈0.2 and the biomass of green picophytoplankton at ≈0.05 Gt C (link to full calculation). We further discuss the uncertainty these caveats may introduce to our estimate of the global mass of plants in the ocean in the section dedicated to plant biomass.

In addition to green algae that are part of the picophytoplankton size group, there are green algae that are larger. To the best our knowledge, we do not know of a good source that constrains the global biomass of green algae in larger size fractions. Therefore, we relied on several lines of evidence to provide a coarse estimate for their global biomass. A recent study (Lopes Dos Santos et al., 2017) quantified the contribution of chlorophyta (green algae) to the total phytoplankton community at all size fractions based on 18S rRNA metabarcoding. Sampling in diverse locations in the ocean, they found that chlorophytes contribute about 10-20% of the total number of photosynthetic reads. Relying on 18S sequence abundance as a proxy for biomass is not a well-established practice, and has various biases. Thus, we also relied on an independent method for estimating the biomass contribution of chlorophytes to the total phytoplankton biomass. A recent report has looked at the phytoplankton community composition in the deep chlorophyll maximum in the North-East Atlantic Ocean (Latasa et al., 2017), based on fluorescent in-situ hybridization (FISH) and flow-cytometry. The average contribution of chlorophyta to the total phytoplankton community reported in this study is ≈15%, which corresponds well with the 18S sequencing data. Thus, our best estimate for the global biomass of chlorophytes is 15% of the total biomass of phytoplankton .

For diatoms, Leblanc et al. (Leblanc et al., 2012) estimates a global biomass of ≈0.3 Gt C (link to full calculation). For *Phaeocystis*, the estimate in Buitenhuis et al. (Buitenhuis et al., 2013) is ≈0.3 Gt C, but it was noted that the data lacks coverage and has a bias to coastal environments (link to full calculation). As stated in Buitenhuis et al. (Buitenhuis et al., 2013), 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 the overestimation of *Phaeocystis* biomass because of sampling bias toward coastal areas.

Combining all the above sources, we estimate the total biomass of phytoplankton to be ≈1.1 Gt C, with the global biomass of bacterial, plant and protist phytoplankton estimated at 0.17 Gt C, 0.17 Gt C and 0.8 Gt C, respectively (link to full calculation).

Top-down approach

Our top-down approach for estimating the total biomass of phytoplankton is based on studies measuring phytoplankton biomass by remote sensing of the reflectance spectrum of the ocean and calculating the phytoplankton-specific absorption coefficients (Behrenfeld et al., 2005; Graff et al., 2015). Best estimates based on this approach put the global biomass of phytoplankton at ≈0.4 Gt C (Behrenfeld et al., 2005; Silsbe et al., 2016). To segregate the total biomass of phytoplankton into bacterial, plant and protists fractions, we relied on our previous estimate that chlorophyta represent ≈15% of the total biomass of phytoplankton, or ≈0.06 Gt C (link to full calculation). In addition, we relied on an independent source for estimating the total biomass of cyanobacteria (Flombaum et al., 2013). Flombaum et al. estimate a total of $\approx 3\times10^{27}$ and $\approx 7\times10^{26}$ cells for *Prochlorococcus* and *Synechococcus*, respectively. To convert cellular abundances into biomass, we relied on estimates for the carbon content of both *Prochlorococcus* and *Synechococcus*. Data from the literature, both from culture data and from indirect measurements in the field, suggest a range of 16-92 fg C cell^{-1} for *Prochlorococcus* and 82-600 fg C cell^{-1} for *Synechococcus* (Buitenhuis et al., 2012b). The data suggests that the carbon content of *Prochlorococcus* and *Synechococcus* is dependent on the ambient illumination conditions, with cells containing lower carbon content at higher illumination levels (Latasa et al., 2017) and stabilizing at ≈50 fg C cell^{-1} for *Prochlorococcus* and ≈100 fg C cell^{-1} for *Synechococcus (Latasa et al., 2017)* when subject to above 1% of the surface photosynthetically active radiation. Based on their global distribution (Flombaum et al., 2013), most *Prochlorococcus* and *Synechococcus* are found in locations in which ambient light is above this threshold, so we used a carbon content of ≈50 fg C cell^{-1} for *Prochlorococcus* and ≈100 fg C cell^{-1} for *Synechococcus*, which correspond to a global estimate of ≈0.2 Gt C. This means that the remaining biomass of phytoplankton, contributed by protists, is ≈0.12 Gt C (link to full calculation).

Combining the two approaches

Each of the two approaches we used has some inherent caveats associated with it. Specifically, the bottom-up approach is largely based on integrating data from various studies that have measured the abundance of phytoplankton using different methods. To convert all these abundance measurements into biomass, the MAREDAT database uses conversion ratios between different measures and carbon mass. These conversion ratios have uncertainty associated with them, and may introduce systematic biases to the global estimate of the biomass of each phytoplankton group. Additionally, there might be a sampling bias of each group to the locations in which it is more dominant. The global biomass of each group is based on the average concentration of biomass across the sampled locations. If most of the samples originate from locations of high abundance, this may lead to an extrapolation of high abundance across the entire ocean, also in locations in which the phytoplankton group is absent, and thus lead to an overestimate.

On the other hand, remote sensing-based estimates of the biomass of phytoplankton may not represent the full scope of the distribution of phytoplankton both in terms of space and in terms of phytoplankton diversity. The reflectance spectrum based on which the biomass of phytoplankton was estimated is not uniformly sensitive to all sizes of phytoplankton (Graff et al., 2015), and may underrepresent specific groups, such as larger types of phytoplankton, like diatoms. Additionally, remote sensing can best measure the biomass of phytoplankton close to the surface of the ocean, whereas in some locations the majority of phytoplankton biomass is found at greater depths in the water column (Latasa et al., 2017).

Thus, extrapolating the biomass of phytoplankton to deeper parts of the water column may be challenging and remote sensing-based estimates might underestimate the global biomass of phytoplankton. As our best estimate of the global biomass of each group of phytoplankton, we used the geometric mean of the estimates based on the top-down and bottom-up approaches. We estimate the global biomass of cyanobacteria, green microalgae and protist phytoplankton at **≈0.2 Gt C**, **≈0.1 Gt C** and **≈0.3 Gt C**, respectively (link to full calculation).

We analyzed the associated uncertainty of the estimate for the total biomass of different groups of phytoplankton, which we report as a fold-change factor from the mean representing a range akin to the 95% confidence interval of the estimate. We use the difference between our bottom-up and top-down approaches to estimate the uncertainty associated with each group. We do not project uncertainty for cyanobacteria, as their biomass is already considered within the global biomass of marine bacteria. This yields a projection of uncertainty of ≈3-fold and ≈6-fold for plant phytoplankton and protist phytoplankton, respectively, as described in the following sections (link to full calculation).

Bacteria and Archaea

Our estimate of the total biomass of marine bacteria and archaea is based on the values reported in (Bar-On et al., 2018). Specifically, we estimate a total of ≈1.3 Gt C of marine bacteria and ≈0.3 Gt C of marine archaea. The estimate of the total biomass of marine bacteria and archaea is based on estimating the total number of bacterial and archaeal cells, and then multiplying them by the characteristic carbon content of a bacterial and archaeal cell in the ocean. A caveat for using this approach is that the methodology for calculating the characteristic carbon content of cells may miss rare cells with a large carbon content. An example of such a case is cyanobacteria. While numerically *Prochlorococcus* and *Synechococcus* represent at most ≈3% of the total number of cells, due to their larger than average size, their biomass contribution is not negligible. When calculating the characteristic carbon content of bacterial and archaeal cells, these rare cells may not be observed, and thus the characteristic carbon content of bacterial and archaeal cells will be underestimated. To address this caveat, even if partially, we added to our estimate of the total biomass of marine bacteria the contribution from cyanobacteria. Additionally, to reflect this added source of uncertainty, we increased our projections for the uncertainty associated with our estimates of the total biomass of marine bacteria and archaea to ≈3-fold.

Plants

We consider in our analysis three broad groups of marine plants - microscopic green algae, which are part of the *Viridiplantae* clade, macroscopic green (part of *Viridiplantae*) and red (part of the *Rhodophyta* clade) algae, and seagrasses, which are flowering plants. We estimated the global biomass of each group separately and then combined the three estimates to arrive at our best estimate for the global mass of marine plants. Overall, we estimate that marine plants have a global biomass of ≈**0.5** Gt C (link to full calculation). By combining the uncertainties associated with the estimates of each group of marine plants, we can project the associated uncertainty of the estimate for the total biomass of marine plants, which we report as a fold-change factor from the mean representing a range akin to the 95% confidence interval of the estimate. We project the uncertainty associated with the estimate of the total biomass of marine plants to be ≈3-fold (link to full calculation). This uncertainty projection does not include the division of

macroalgal biomass between plants and protists, but this type of uncertainty is not expected to influence the overall uncertainty significantly beyond the uncertainty associated with the estimate of the total biomass of marine plants, as its maximal influence increases or decreases the biomass of plants by 0.2 Gt C, which, compared to our best estimate of ≈0.5 Gt C, is on a smaller scale than the overall uncertainty associated with our estimate of the total biomass of marine plants.

Green microalgae

As discussed in the phytoplankton section of the SI, our best estimate for the global biomass of green microalgae is ≈0.1 Gt C (link to full calculation). In the section estimating the biomass of phytoplankton, we projected the uncertainty of this estimate to be ≈3-fold (link to full calculation). As discussed in the phytoplankton section, this estimate has a large uncertainty associated with it, as the number of studies it is based on is limited. We note that even if the uncertainty associated with the estimate of the biomass of green microalgae is increased to 10-fold, to take into account the additional uncertainty due to the lack of data, it would not have a large influence on the total uncertainty associated with our estimate of the total biomass of marine plants due their overall small biomass.

Macroalgae

Macroalgae consist of a diverse group of multicellular algae, which includes both green (*Chlorophyta*) and red (*Rhodophyta*) algae belonging to the plant (*Archaoplastida*) clade, as well as brown algae (*Phaeophytes*), which are part of the protist clades. Macroalgae can be both benthic (attached to the seafloor) or planktonic (e.g., the common brown algae genus *Sargassum*). We first estimated the total biomass of benthic macroalgae, i.e., green, red and brown variants. We note that holoplanktonic algae (algae that spend their entire life cycle in open waters) are brown algae (which are protists), so we included them in the estimate of the global biomass of marine protists.
Our estimate for the global mass of benthic macroalgae is based on estimates for the global area in which macroalgae are distributed, together with estimates of the biomass density of macroalgae per unit area. The best available evidence (Duarte, 2017) suggests that the global range of macroalgae is somewhere between 1.4×10^{12} m^2 and 5.7×10^{12} m^2. The lower number is based on the global extent of macrophytes (seagrasses and macroalgae) stated in Whittaker and Likens (Whittaker and Likens, 1973) minus the possible extent of seagrasses from Duarte and Chiscano (Duarte and Chiscano, 1999). The higher value is based on the depth at which sufficient light is available to allow for benthic macroalgal growth (Gattuso et al., 2006). As a best estimate, we used the geometric mean of these two bounds, which is $\approx2.8\times10^{12}$ m^2. For the characteristic biomass density of macroalgae, we used data from Cebrian et al. (Cebrian et al., 2009), which report an average of ≈180 g C m^{-2} and a geometric mean of ≈120 g C m^{-2}. As our best estimate for the biomass density of macroalgae, we used the geometric mean of the average biomass density and the geometric mean biomass density (see our section regarding estimators of the mean), which is ≈150 g C m^{-2}. By applying this biomass density across our best estimate for the global range of macroalgae, we arrived at our best estimate for the global biomass of macroalgae, which is ≈0.4 Gt C (link to full calculation).

We could not find adequate data to estimate how the global biomass of macroalgae is distributed between plants (green and red algae) and protists. The current best estimate for the global biomass of kelp forests, which are protist macroalgae, is ≈0.02 Gt C (Laffoley and Grimsditch, 2009), which leaves us with ≈0.4

Gt C to be divided between plants and protists. As we do not have any knowledge as to the specific distribution of macroalgal biomass, we divided it equally between plant biomass and protist biomass.

We now present our analysis for the associated uncertainty of the estimate for the total biomass of macroalgae, which we report as a fold-change factor from the mean representing a range akin to the 95% confidence interval of the estimate. For the biomass density of macroalgae, we used the variance between reported values in Cebrian et al. (Cebrian et al., 2009) to project an uncertainty of ≈1.7-fold, associated with our estimate of the characteristic biomass density of macroalgae. For the area covered by macroalgae, we used the maximum and minimum estimates of the global area to project an uncertainty of ≈2-fold, associated with our estimate. By combining the uncertainties associated with both the biomass density per unit area and the global distribution of macroalgae, we project an uncertainty of ≈2-fold, associated with our estimate of the biomass of macroalgae (link to full calculation).
When data is ample, the uncertainty around the geometric mean will be low (as we base our uncertainty on the SE). Nevertheless, this type of uncertainty does not consider the possibility that the distribution of values in the sample data does not represent the natural environment faithfully, which is probably common in measurements that are sparse and biased. As the uncertainty projection for macroalgae seems low compared to the amount of data available, we decided to also analyze the uncertainty by employing a procedure to take into account the possible effect of bias in our samples. We generate an additional multiplicative uncertainty based on the SD and not on the SE in log-space. We consider the SE-based multiplicative uncertainty as an underestimate of the actual uncertainty and the SD-based multiplicative uncertainty as an overestimate of the actual uncertainty (because it does not include the decrease in uncertainty due to averaging). As our measure of uncertainty, we use the geometric mean of the SE-based multiplicative uncertainty and the SD-based multiplicative uncertainty. While this is not a standard statistical procedure, we consider it to be a reasonable compromise for deriving a robust uncertainty estimate. This procedure increases the uncertainty associated with the estimate of the total biomass of macroalgae to ≈5-fold (link to full calculation). It does not significantly change the total uncertainty associated with the estimate of the total biomass of marine plants.

Seagrasses

For seagrass, we relied on the estimate by Fourqurean et al. (Fourqurean et al., 2012) for the average biomass density of seagrasses. Fourqurean et al. report an average biomass density of ≈250 g C m^{-2} and a median biomass density of ≈100 g C m^{-2}. As our best estimate, we used the geometric mean of the average density and the median density (see our section regarding estimators of the mean), which is ≈160 g C m^{-2}. For the global range of seagrasses, we used two estimates, one based on *in-situ* observed occurrence of seagrasses (Green et al., 2003; Jayathilake and Costello, 2018), which is ≈0.8×10^{12} m^2, and the other based on modeling the global distribution of seagrasses based on environmental parameters, which is ≈1.6×10^{12} m^2 (Jayathilake and Costello, 2018). As our best estimate, we used the geometric mean of the two estimates, which is ≈10^{12} m^2. Applying the characteristic biomass density of seagrasses across the range of seagrasses, we arrived at our estimate for the total biomass of seagrasses, ≈0.2 Gt C (link to full calculation).

We analyzed the associated uncertainty of the estimate for the total biomass of seagrasses, which we report as a fold-change factor from the mean, representing a range akin to the 95% confidence interval of the estimate. We project the uncertainty associated with the estimate of the biomass density of seagrasses to be ≈2.5-fold. We project the uncertainty associated with the estimate of the global area covered by seagrasses to be ≈2-fold. By combining these two uncertainties, we project the uncertainty associated with our estimate of the global biomass of seagrasses to be ≈3-fold (link to full calculation).

Animals

We considered in our analysis several broad groups of marine animals - annelids, nematodes, arthropods, chordates, cnidarians, molluscs and other benthic phyla. We estimated the global biomass of each group separately and then combined the estimates to arrive at our best estimate for the global mass of marine plants. Overall, we estimate that marine plants have a global biomass of **≈2 Gt C** (link to full calculation). By combining the uncertainties associated with the estimates of each group of marine plants, we can project the associated uncertainty of the estimate for the total biomass of marine animals, which we report as a fold-change factor from the mean representing a range akin to the 95% confidence interval of the estimate. We project the uncertainty associated with the estimate of the total biomass of marine animals to be ≈3-fold. This uncertainty projection does not include the additional possible contribution from benthic phyla in hotspots such as seamounts and submarine canyons, which we found to be less than 1 Gt C.

Chordates

The main groups of marine chordates we considered in our analysis are fish, marine mammals and tunicates. We estimated the global biomass of each group separately and then combined the estimates of each group to arrive at our best estimate of the total biomass of chordates.

Fish

The global fish biomass was historically estimated based on trawling surveys and primary productivity coupled models of trophic transfer efficiencies (Gjøsaeter et al., 1980; Lam and Pauly, 2005; Wilson et al., 2009). These estimates put the global fish biomass at around ≈0.3 Gt C (2 Gt fresh weight), with half of the fish biomass contributed by mesopelagic fish (fish that live at 200-1000 m depth (Wilson et al., 2009), which corresponds well with trawling-based estimates of mesopelagic fish biomass (Lam and Pauly, 2005). Estimating mesopelagic fish biomass using trawl data could lead to underestimation, because fish can avoid the trawls, as demonstrated by Kaartvedt et al. (Kaartvedt et al., 2012). An independent method for estimating the biomass of mesopelagic fish uses sonar to record echoes backscattered by gas-filled swim bladders of mesopelagic fish. In 2010, as part of the Malaspina expedition, measurements of backscatter from sonar echosounder were collected from various locations across the globe. A recent study (Irigoien et al., 2014) 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 throughout the oceans. An estimate of the mesopelagic biomass is based on the measurement of acoustic backscatter from the deep scattering layer, combined with calibrations of the strength of reflection of acoustic signals from a single mesopelagic fish, termed target strength. The study by Irigoien et al.(Irigoien et al., 2014) measured the total scattering along the course of the campaign and found a correlation between this scattering strength and the local net primary productivity (NPP). Then,

using estimates of NPP across the ocean, Irigoien et al. (Irigoien et al., 2014) extrapolated the total scatter across the entire ocean. The specific parameters of the correlation between NPP and scattering strength are given in detail in (Irigoien et al., 2014). Irigoien et al. converted 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. Estimating the biomass of mesopelagic fish using acoustic measurements has many caveats. These include possible resonance of the sonar beam, which leads to an overestimate of the backscatter, uncertainty regarding the fraction of fish with a gas-filled swim bladder, and the fraction of the total backscatter contributed by mesopelagic fish. To take these issues into account, later studies using an independent set of acoustic measurements from the deep scattering layer generated a new estimate for the global biomass of mesopelagic fish (Proud et al., 2019). This updated estimate suggests that the biomass of mesopelagic fish could vary significantly, depending on the assumptions made regarding the abovementioned caveats. Proud et al. generated three different estimates for the global biomass of mesopelagic fish, each with a different assumption regarding the fraction of fish with a gas-filled swim bladder in the population. We relied on the geometric mean of the three different estimates, ≈1.4 Gt C, as our best estimate of mesopelagic fish biomass, based on acoustic measurements (link to full calculation). This estimate does not take into account the issue of the fraction of the total acoustic backscatter attributed to mesopelagic fish. In addition to mesopelagic fish, siphonophores, which are a colonial form of hydrozoans, can also contribute significantly to the global backscatter (Proud et al., 2019), but currently there is very limited data to constrain their contribution to the global acoustic backscatter. Thus, the estimate based on acoustic measurements is likely an overestimate. Our best estimate of the global biomass of mesopelagic fish is the geometric mean of the estimate based on trawling (which is likely an underestimate) and the estimate based on acoustic measurements (which is likely an overestimate), which is ≈0.45 Gt C. We add to this geometric mean the contribution from other groups of fish, ≈0.15 Gt C (Wilson et al., 2009), to yield a final estimate of **≈0.6 Gt C** (link to full calculation).

We now present our analysis of 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 the 95% confidence interval of the estimate. We start by analyzing the associated uncertainty of the estimate of the biomass of mesopelagic fish. We first considered the uncertainty associated with the estimate based on acoustic measurements. In their study, Proud et al. report an uncertainty of ≈2-fold for the 25%-75% percentiles distribution range. Assuming the distribution of estimates is close to log-normal, this means that the 95% confidence interval of their estimate is ≈4-fold. To consider also the uncertainty associated with the fraction of the backscatter contributed by mesopelagic fish, we calculated the uncertainty in the estimate of the total mesopelagic fish biomass between the two independent methods, which is ≈10-fold (link to full calculation). 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 relied on estimates by Wilson et al., which do not report an uncertainty range for the biomass of non-mesopelagic fish. A later study (Jennings and Collingridge, 2015) gave an estimate for the total biomass of marine animals with a body weight between 1 g and 1000 kg, based on ecological models. Jennings et al. report a median estimate of ≈0.75 Gt C for all marine animal biomass, which includes both mesopelagic and non-mesopelagic fish biomass, as well as other marine animals. 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 7-fold associated with the estimate of the global biomass of fish (link to full calculation).

Other animal phyla

In (Bar-On et al., 2018) we used a bottom-up approach that 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 (similar to (Bar-On et al., 2018)). 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 relied on the bounds reported in (Bar-On et al., 2018) for the total contribution for these remaining phyla, which are, for the most part, benthic organisms. In (Bar-On et al., 2018), we estimated 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 are underrepresented. Although estimates for the contribution of each animal group 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 in the main text of the work.

Other than benthic animal biomass, some groups of planktonic animals may be underrepresented. More specifically, delicate animals which may get degraded in the sampling process could be underestimated. One such group is siphonophores, which are a colonial form of hydrozoans. Their biomass is currently considered within our estimate for the total biomass of macrozooplankton, but because the estimate relies on sampling by nets, it could significantly underestimate siphonophore biomass. Recent studies have indicated that siphonophores can also contribute significantly to the measurements of marine acoustic backscatter used to estimate the biomass of the deep scattering layer, which is composed of many organisms, including mesopelagic fish (Proud et al., 2019). Currently, however, there is very limited data to constrain their contribution to the marine acoustic backscatter.

Protists

We consider several dominant groups of marine protists, broadly categorized into several groups - planktonic protists, protist macroalgae, microphytobenthos and particle-attached protists. By combining our estimates for the biomass of all the above groups, we estimate the total biomass of protists in the ocean at **≈2 Gt C,** as detailed below (link to full calculation). By combining the uncertainties associated with the estimates of each group of marine protists, we can project the associated uncertainty of the estimate for the total biomass of marine protists, which we report as a fold-change factor from the mean representing a range akin to the 95% confidence interval of the estimate. We project the uncertainty associated with the estimate of the total biomass of marine protists to be ≈3-fold (link to full calculation). This uncertainty does not include the division of macroalgal biomass between plants and protists, and the uncertainty associated with the estimate of the biomass of heterotrophic pico-nanoplankton. However, this type of uncertainty is not expected to influence the overall uncertainty significantly.

Planktonic protists

In the section discussing the biomass of phytoplankton above, we estimated the global biomass of protist phytoplankton at ≈0.3 Gt C. In addition to autotrophic protists, we consider in this section heterotrophic planktonic protists, which belong to the following groups: pico- and nanoplankton, microzooplankton (Buitenhuis et al., 2013) (defined not to include copepod biomass), deep sea protists and *Rhizaria*. 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.

Some of our estimates for the biomass of planktonic protists are based on the MAREDAT database (Buitenhuis et al., 2013). Members of the MAREDAT initiative used this database to estimate the global biomass of many plankton functional types, using a characteristic biomass concentration for each depth (either the median or average of the values in the database), and applying it 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 to the right 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 (which are actually the median and mean values of the distribution reported in the database) as our best estimate, which we employ in order to increase the 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.

Protists in the picoplankton to nanoplankton size range (0.8-5 μm in diameter) include both autotrophic and heterotrophic organisms. In the section discussing the global biomass of phytoplankton, we estimated the biomass of picoeukaryotic autotrophs as ≈20% of the total biomass of phytoplankton, or ≈0.12 Gt C. As we could not find a reliable resource for estimating the biomass of heterotrophic pico- and nanoplankton, we used a recent global 18S ribosomal DNA sequencing effort that was part of the *Tara* Oceans campaign (de Vargas et al., 2015). 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 ≈0.25 Gt C of heterotrophic pico-nanoplankton protists (link to full calculation). Relying on 18S sequence abundance as a proxy for biomass is not a well-established practice, and has various biases, but for lack of an alternative for the estimate, we chose to use it. We note that in any case this plays a minor role in our analysis and thus will not affect any of the major conclusions of our study.

In addition to picoplankton, protists are also abundant in larger size fractions of plankton. Protists in the microzooplankton size fraction (5-200 µm in diameter), which include ciliates, dinoflagellates, flagellates and amoeba, are important grazers of phytoplankton (Buitenhuis et al., 2010). Our best estimate of the global biomass of protist microzooplankton, which is based on the MAREDAT database (Buitenhuis et al., 2013) is ≈0.6 Gt C (link to full calculation). We note that in the MAREDAT database, the microzooplankton category was defined not to include contributions from copepods, and is thus dominated by protist biomass.

Some protists are even larger than the size fraction of microzooplankton. An important group of large protists are the *Rhizaria* super-group. A recent paper by Biard et al. (Biard et al., 2016) 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 the median biomass concentrations at each depth layer across the entire volume of water at that layer to generate a global estimate. We relied on the measurements in the entire range 0-500 m, and calculated the arithmetic and geometric means of the data. Our best estimate for the characteristic density of *Rhizaria* based on the data in Biard et al. is ≈1 g C m^{-2}. By applying this density across the entire area of the ocean, we estimate the biomass of *Rhizaria* in the top 500 meters of the ocean at ≈0.5 Gt C (link to full calculation).

Most of our estimates for the different types of planktonic protists do not include deep sea protists (depth > 200 m for phytoplankton and microzooplankton, and > 500 m for *Rhizaria*). For phytoplankton, a recent paper (Agusti et al., 2015) 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 the omission of deep sea phytoplankton biomass. Similarly for heterotrophic protists, cell abundance in the deep ocean is much lower, with estimates for their total biomass of ≈0.1 Gt C (Pernice et al., 2015). For *Rhizaria*, because of lack of sampling below 500 meters, our estimate might be an underestimate of actual *Rhizaria* biomass.

We now present our analysis of the associated uncertainty of the estimate for the total biomass of planktonic protists, which we report as a fold-change factor from the mean representing a range akin to the 95% confidence interval of the estimate. Our estimate of the biomass of planktonic protists is the sum of our estimates for the biomass of protist phytoplankton, microzooplankton and *Rhizaria*. We project the uncertainty associated with the estimates of each group and then combine these uncertainties to arrive at our best projection of the uncertainty associated with our estimate of the biomass of planktonic protists. For protist phytoplankton, we used the uncertainty we derived in the section above related to phytoplankton biomass, which is ≈6-fold (link to full calculation). For microzooplankton, the estimate is based on 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 project a crude uncertainty of ≈10-fold associated with our estimate of the biomass of microzooplankton. For *Rhizaria*, we relied on the variance in between different samples across the ocean

to project an uncertainty of 1.2-fold associated with the estimate of the biomass of *Rhizaria* (link to full calculation).
When data is ample, the uncertainty around the geometric mean will be low (as we base our uncertainty on the SE). Nevertheless, this type of uncertainty does not consider the possibility that the distribution of values in the sample data does not represent the natural environment faithfully, which is probably common in measurements that are sparse and biased. As the uncertainty projection for *Rhizaria* seems low compared to the amount of data available we decided to also analyze the uncertainty by employing a procedure to take into account the possible effect of bias in our samples. We generate an additional multiplicative uncertainty based on the SD and not on the SE in log-space. We consider the SE-based multiplicative uncertainty as an underestimate of the actual uncertainty and the SD-based multiplicative uncertainty as an overestimate of the actual uncertainty (because it does not include the decrease in uncertainty due to averaging). As our measure of uncertainty, we use the geometric mean of the SE-based multiplicative uncertainty and the SD-based multiplicative uncertainty. While this is not a standard statistical procedure, we consider it to be a reasonable compromise for deriving a robust uncertainty estimate. This procedure increases the uncertainty associated with the estimate of the total biomass of *Rhizaria* to ≈8-fold (link to full calculation). It does not significantly change the total uncertainty associated with the estimate of the biomass of marine protists.

Macroalgae

There are various types of macroalgal protists, which we consider in our analysis. One dominant and well-known group of protist macroalgae is kelp. Estimates of the global biomass of kelp growing on rocky substrates put their global standing crop at 0.02 Gt C (Laffoley and Grimsditch, 2009). A similar value can be calculated based on the annual productivity of kelp given by De vooys (De Vooys, 1979), assuming one turnover of the standing crop each year (Smith, 1981).
In addition to kelp, we analyze other species of protist macroalgae. For our estimate for those protist macroalgae, we relied on the estimate of 0.2 Gt C derived in the plant macroalgae section (link to full calculation). In addition to benthic macroalgae, there are some forms of macroalgae that have a holopelagic lifecycle, meaning that they develop solely in the water column. There are two dominant species that have this life form, both from the genus *Sargassum*. These species are distributed in the Atlantic ocean. Previous estimates have suggested their global wet weight biomass is ≈0.01 Gt (Wang et al., 2018). A recent paper used remote sensing calibrated with field measurements to estimate the global biomass density of *Sargassum* in the North Atlantic. The higher bound of their estimates for the global biomass density of *Sargassum*, even when including the possible contribution of *Sargassum* biomass in patches that are below the detection limit of remote sensing, is ≈0.07 g C m^{-2} (Wang et al., 2018), measured in July 2015, which is around the peak season for *Sargassum* biomass (Wang and Hu, 2017). By applying this density across the entire area of the North Atlantic ocean, which is ≈4×10^{13} m^2, we arrive at an estimate of ≈0.003 Gt C, which is negligible compared to other forms of macroalgae. Recent reports have also found *Sargassum* in the northern shores of Brazil, and have suggested that it does not originate from the North Atlantic (Sissini et al., 2016). Even if we apply the biomass densities from the North Atlantic to the entire Atlantic ocean, we arrive at an estimate of ≈0.007 Gt C, which is still negligible.
In other oceans, such as near China or Japan, floating forms of *Sargassum* originate from benthic algae, and thus are incorporated within our estimates for the global biomass of benthic macroalgae.

We project the same uncertainty associated with our estimate of the biomass of benthic protist macroalgae as the uncertainty we projected for all benthic macroalgae in the plant section above, which is ≈2-fold (link to full calculation).

Benthic microalgae

For microphytobenthos, existing data on their productivity (Charpy Roubaud and Sournia, 1990) puts their global productivity at ≈0.3 Gt C yr^{-1}. Existing data on their turnover puts it at a similar range as phytoplankton, around 0.1 d^{-1} (Cebrian, 1999). Using these values, we arrive at an estimate of their global biomass of ≈0.01 Gt C. With an independent approach based on measured biomass densities per unit area (Cebrian et al., 2009), we arrive at an average density of ≈4 g C m^{-2}. Applying this average biomass density across the entire range of benthic microalgae (which we assume to be similar to that of benthic macroalgae as in (Charpy Roubaud and Sournia, 1990)), we similarly arrive at an estimate of ≈0.01 Gt C. In both analyses, we arrive at similar values, which are negligible with respect to other benthic autotroph biomass or microalgal biomass.

Particle-attached protists

Protists can also be found attached to particulate organic matter. Such particulate organic carbon is usually separated into two broad groups - microaggregates, whose diameter is smaller than 500 μm, and macroaggregates (also known as marine snow), whose diameter is larger than 500 μm. We first considered the abundance of protist on macroaggregates, as more data is available regarding this size fraction of particulate organic carbon. To estimate the biomass of particle-attached protists in macroaggregates (marine snow), we relied on several studies that have measured the relative number of particle-attached protists and prokaryotes at the epipelagic (Herndl, 1988; Turley and Mackie, 1994), mesopelagic (Turley and Mackie, 1994) and bathypelagic (Bochdansky et al., 2017) layers. For each study, we calculated the characteristic ratio between the number of protist and prokaryote cells. To estimate the biomass of particle-attached protists relative to prokaryotes, we used estimates on the carbon content of protists in the epipelagic, mesopelagic and bathypelagic layers (Herndl, 1988; Pernice et al., 2015). We compared this carbon content to our best estimate of the carbon content of particle-attached prokaryotes (see marine bacteria and archaea section). We generated 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 ≈0.12 Gt C of particle-attached prokaryotes in marine snow, thus leading to an estimate of ≈0.12 Gt C of particle-attached protists in marine snow. We could find very little data on the abundance of protists in microaggregates, but according to one study using ribosomal DNA copies, eukaryotes (which include both protists and other eukaryotes) account for ≈20% of the microbial community on microaggregates (Fontanez et al., 2015). We estimate ≈0.15 Gt C of particle-attached prokaryotes in microaggregates, thus leading to a maximal estimate of ≈0.03 Gt C of particle-attached protists in microaggregates. Relying on ribosomal DNA copy abundance as a proxy for biomass is not a well-established practice, and has various biases, but for lack of an alternative for the estimate, we chose to use it. Combining our estimates for the biomass of particle-attached protists in marine snow (macroaggregates) and in microaggregates, we estimate a total of ≈0.15 Gt C of particle-attached protists (link to full calculation).

We now present our analysis of the associated uncertainty of the estimate for the total biomass of particle-attached protists, which we report as a fold-change factor from the mean representing a range akin to the 95% confidence interval of the estimate. For the biomass of particle-attached bacteria and archaea, we project an uncertainty of ≈5-fold associated with our estimate (Bar-On et al., 2018). We estimate the biomass of particle-attached protists by combining the estimate for the biomass of particle-attached prokaryotes and estimates for the ratio between the biomass of protists and prokaryotes in particulate organic carbon. The data on which we base our estimate for the ratio between the biomass of protists and prokaryotes in particulate organic carbon is very scarce. Due to the presence of many uncertainty sources, most of which are hard to quantify, we could not rely on reported uncertainties within or across studies, as those would give gross underestimates, and chose instead to project a 10-fold uncertainty associated with our estimate of the biomass of particle-attached protists.

Fungi

To estimate the total biomass of marine fungi, we considered several environments in which fungi reside: planktonic fungi and particle-attached fungi. We used the estimate for the total biomass of planktonic fungi from (Bar-On et al., 2018) of ≈0.15 Gt C. For particle-attached fungi, we distinguish between two types of particles - microaggregates, whose diameter is smaller than 500 μm, and macroaggregates (also known as marine snow), whose diameter is larger than 500 μm. Regarding macroaggregates, a recent study by Bochdansky et al.(Bochdansky et al., 2017) has measured the relative biomass of particle-attached fungi and prokaryotes in the bathypelagic layer. We used 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 applied the ratio measured in the bathypelagic layer across the entire volume of the ocean. We arrived at an estimate ≈0.12 Gt C of particle-attached prokaryotes on macroaggregates, so we estimate ≈0.08 Gt C particle-attached fungi (link to full calculation). For microaggregates, we could find very little data, but according to one study using ribosomal DNA copies, eukaryotes (which include both fungi and other eukaryotes) account for ≈20% of the microbial community of microaggregates (Fontanez et al., 2015). We estimate ≈0.15 Gt C of particle-attached prokaryotes in microaggregates, thus leading to a maximal estimate of ≈0.03 Gt C of particle-attached protists in microaggregates. Relying on ribosomal DNA copy abundance as a proxy for biomass is not a well-established practice, and has various biases, but for lack of an alternative for the estimate, we chose to use it. By combining our estimates for the biomass of particle-attached protists in marine snow (macroaggregates) and microaggregates, we estimate a total of ≈0.1 Gt C of particle-attached protists (link to full calculation).

Overall, by combining data on both planktonic marine fungi and particle-attached fungi, we estimate that the biomass of marine fungi is **≈0.3 Gt C** (link to full calculation). A general caveat associated with estimating the biomass of marine fungi is that this group has been understudied, 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.

Viruses

To estimate the biomass of viruses in the ocean, we first estimated the total number of virus particles in the ocean, as well as the characteristic carbon content of a single virus. We then multiplied the total number of viruses by the characteristic carbon content of a single virus to arrive at our estimate for the total biomass of marine viruses. To estimate the total number of viruses in the ocean, we relied on data from two recent studies - the first is Wigington et al. (Wigington et al., 2016), which provided ≈6000 samples of virus concentration in the ocean, and the second is Lara et al. (Lara et al., 2017), which included data from the global Malaspina campaign. To transform this data into a robust estimate of the global biomass of marine viruses, we used three different methodologies to extrapolate the total number of marine viruses, and applied each on both datasets, generating a total of six different estimates. The first two methodologies we used are based on binning the data along the depth of the ocean into 10 quantiles. For each bin, we calculated either the mean concentration of viruses or the geometric mean of the concentration of viruses. We then multiplied the mean concentration of viruses (or geometric mean) 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. The third methodology is based on fitting a linear regression model to predict the log of concentration of viruses based on the log of the depth of the sample, and then using this model to extrapolate the concentration of viruses at all ocean depths and integrating over the entire volume of the ocean. As our best estimate for the total number of viruses in the ocean, we used the geometric mean of the six different estimates, which is $\approx 1.5 \times 10^{30}$ (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 (Brum et al., 2015) (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 virus biomass. Therefore, we assumed bacteriophages dominate the biomass of marine viruses, and we used the characteristic carbon content of a single bacteriophage to estimate the total biomass of marine viruses.
To estimate the characteristic carbon content of a single bacteriophage, we relied on a biophysical model of the elemental composition of a virion (Jover et al., 2014). By plugging into the formulas detailed in (Jover et al., 2014) a characteristic radius of 25 nm (based on a characteristic diameter of 50 nm; (Brum et al., 2015)), we estimate that a characteristic bacteriophage contains about ≈0.02 fg C (link to full calculation). Previous reports have suggested a carbon content range of 0.05-0.2 fg C per virion (Jover et al., 2014), which may reflect higher estimates for the size of marine viruses (a capsid radius of ≈50 nm could translate into 0.2 fg C). However, recent data from Brum et al., suggests marine viruses are on average smaller and thus we used a lower estimate for their carbon content. To estimate the total biomass of viruses, we multiplied the characteristic carbon content of ≈0.02 fg C by our estimate for the total number of viruses, $\approx 1.5 \times 10^{30}$, and arrived at a total biomass of **≈0.03 Gt C** (link to full calculation).

We now present our analysis of the associated uncertainty of the estimate for the total biomass of marine viruses, which we report as a fold-change factor from the mean representing a range akin to the 95%

confidence interval of the estimate. In this analysis, we considered the following factors: First, we assessed the uncertainty associated with the estimate of the total number of marine viruses. To arrive at our projection of the uncertainty associated with our estimate of the total number of marine viruses, we relied on the variation between the six different estimates we generated for the total number of marine viruses. As these estimates are likely to be somewhat dependent on each other, we did not use the standard error of the mean to quantify the variation between the estimates, but rather the standard deviation. Our best projection for the uncertainty associated with our estimate of the total number of marine viruses is ≈1.5-fold (link to full calculation).

To quantify the uncertainty associated with the size range of viruses, we used data from Brum et al.(Brum et al., 2015) on marine viruses. The uncertainty of the average diameter of marine viruses is less than 2-fold (link to full calculation). We propagated the uncertainty in the diameter of bacteriophages into our estimate for the carbon content of a single bacteriophage, the result of which is an uncertainty of ≈2-fold (link to full calculation). We combined the uncertainty in our estimate of the radius of bacteriophages with the uncertainty concerning the parameters of the biophysical model introduced into the estimate of the carbon content of bacteriophages. The uncertainty of the parameters of the model are detailed in Jover et al. (Jover et al., 2014). Overall, we project that an uncertainty of ≈2-fold is 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 estimated the same carbon content for both archaeal and bacterial viruses, even though archaeal viruses are highly diverse (Prangishvili et al., 2017). Nevertheless, the relative biomass of bacteria is higher than that of archaea, and thus it is likely that bacterial viruses are more abundant than archaeal viruses.

When combining the uncertainties associated with our estimates of the total number of marine viruses and the characteristic carbon content of a marine virus, we project an uncertainty of ≈2.5-fold associated with our estimate of the total biomass of marine viruses (link to full calculation).

References

Agusti, S., González-Gordillo, J.I., Vaqué, D., Estrada, M., Cerezo, M.I., Salazar, G., Gasol, J.M., and Duarte, C.M. (2015). Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. Nat. Commun. *6*, 7608.

Bar-On, Y.M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. Proc. Natl. Acad. Sci. U. S. A. *115*, 6506–6511.

Behrenfeld, M.J., Boss, E., Siegel, D.A., and Shea, D.M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space: PHYTOPLANKTON GROWTH RATES AND OCEAN PRODUCTIVITY. Global Biogeochem. Cycles *19*, 57.

Biard, T., Tristan, B., Lars, S., Marc, P., Nicolas, M., Pieter, V., Helena, H., Gabriel, G., Lionel, G., Rainer, K., et al. (2016). In situ imaging reveals the biomass of giant protists in the global ocean. Nature *532*, 504–507.

Bochdansky, A.B., Clouse, M.A., and Herndl, G.J. (2017). Eukaryotic microbes, principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. ISME J. *11*, 362–373.

Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulcier, G., Acinas, S.G., Alberti, A., Chaffron, S.,

Cruaud, C., de Vargas, C., Gasol, J.M., et al. (2015). Patterns and ecological drivers of ocean viral communities. Science *348*, 1261498–1261498.

Buitenhuis, E.T., Rivkin, R.B., Sévrine, S., and Le Quéré, C. (2010). Biogeochemical fluxes through microzooplankton. Global Biogeochem. Cycles *24*.

Buitenhuis, E.T., Li, W.K.W., Vaulot, D., Lomas, M.W., Landry, M.R., Partensky, F., Karl, D.M., Ulloa, O., Campbell, L., Jacquet, S., et al. (2012a). Picophytoplankton biomass distribution in the global ocean. Earth System Science Data *4*, 37–46.

Buitenhuis, E.T., Li, W., and Vaulot, D. (2012b). Picophytoplankton biomass distribution in the global ocean. Earth Syst. Monit.

Buitenhuis, E.T., Vogt, M., Moriarty, R., Bednaršek, N., Doney, S.C., Leblanc, K., Le Quéré, C., Luo, Y.-W., O'Brien, C., O'Brien, T., et al. (2013). MAREDAT: towards a world atlas of MARine Ecosystem DATa. Earth Syst. Sci. Data *5*, 227–239.

Cebrian, J. (1999). Patterns in the Fate of Production in Plant Communities. Am. Nat. *154*, 449–468.

Cebrian, J., Shurin, J.B., Borer, E.T., Cardinale, B.J., Ngai, J.T., Smith, M.D., and Fagan, W.F. (2009). Producer nutritional quality controls ecosystem trophic structure. PLoS One *4*, e4929.

Charpy Roubaud, C., and Sournia, A. (1990). Microalgal growth : inputs and losses, practical approches.

De Vooys, C. (1979). SCOPE 13--The Global Carbon Cycle.

Duarte, C.M. (2017). Reviews and syntheses: Hidden forests, the role of vegetated coastal habitats in the ocean carbon budget. Biogeosciences *14*, 301–310.

Duarte, C.M., and Chiscano, C.L. (1999). Seagrass biomass and production: a reassessment. Aquat. Bot. *65*, 159–174.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K.W., Lomas, M.W., Veneziano, D., et al. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. Proc. Natl. Acad. Sci. U. S. A. *110*, 9824–9829.

Fontanez, K.M., Eppley, J.M., Samo, T.J., Karl, D.M., and DeLong, E.F. (2015). Microbial community structure and function on sinking particles in the North Pacific Subtropical Gyre. Front. Microbiol. *6*, 469.

Fourqurean, J.W., Duarte, C.M., Kennedy, H., Marbà, N., Holmer, M., Mateo, M.A., Apostolaki, E.T., Kendrick, G.A., Krause-Jensen, D., McGlathery, K.J., et al. (2012). Seagrass ecosystems as a globally significant carbon stock. Nat. Geosci. *5*, 505.

Gattuso, J.-P., Gentili, B., Duarte, C. m., Kleypas, J. a., Middelburg, J. j., and Antoine, D. (2006). Light availability in the coastal ocean: impact on the distribution of benthic photosynthetic organisms and their contribution to primary production. Biogeosciences *3*, 489–513.

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

Graff, J.R., Westberry, T.K., Milligan, A.J., Brown, M.B., Dall'Olmo, G., van Dongen-Vogels, V., Reifel,

K.M., and Behrenfeld, M.J. (2015). Analytical phytoplankton carbon measurements spanning diverse ecosystems. Deep Sea Research Part I: Oceanographic Research Papers *102*, 16–25.

Green, E.P., Short, F.T., and . . F. (2003). World Atlas of Seagrasses (University of California Press).

Herndl, G.J. (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*, 265–275.

Irigoien, X., Xabier, I., Klevjer, T.A., Røstad, A., Martinez, U., Boyra, G., Acuña, J.L., Bode, A., Echevarria, F., Gonzalez-Gordillo, J.I., et al. (2014). Large mesopelagic fishes biomass and trophic efficiency in the open ocean. Nat. Commun. *5*.

Jayathilake, D.R.M., and Costello, M.J. (2018). A modelled global distribution of the seagrass biome. Biol. Conserv. *226*, 120–126.

Jennings, S., and 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*, e0133794.

Jover, L.F., Effler, T.C., Buchan, A., Wilhelm, S.W., and Weitz, J.S. (2014). The elemental composition of virus particles: implications for marine biogeochemical cycles. Nat. Rev. Microbiol. *12*, 519–528.

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

Laffoley, D., and Grimsditch, G.D. (2009). The Management of Natural Coastal Carbon Sinks (IUCN).

Lam, V., and Pauly, D. (2005). Mapping the global biomass of mesopelagic fishes. Sea Around Us Proj. Newsl. *30*.

Lara, E., Vaqué, D., Sà, E.L., Boras, J.A., Gomes, A., Borrull, E., Díez-Vives, C., Teira, E., Pernice, M.C., Garcia, F.C., et al. (2017). Unveiling the role and life strategies of viruses from the surface to the dark ocean. Sci Adv *3*, e1602565.

Latasa, M., Cabello, A.M., Morán, X.A.G., Massana, R., and Scharek, R. (2017). Distribution of phytoplankton groups within the deep chlorophyll maximum. Limnol. Oceanogr. *62*, 665–685.

Leblanc, K., Arístegui, J., Armand, L., Assmy, P., Beker, B., Bode, A., Breton, E., Cornet, V., Gibson, J., M.-P., G., et al. (2012). A global diatom database – abundance, biovolume and biomass in the world ocean. Earth System Science Data *4*, 149–165.

Limardo, A.J., Sudek, S., Choi, C.J., Poirier, C., Rii, Y.M., Blum, M., Roth, R., Goodenough, U., Church, M.J., and Worden, A.Z. (2017). Quantitative biogeography of picoprasinophytes establishes ecotype distributions and significant contributions to marine phytoplankton. Environ. Microbiol. *19*, 3219–3234.

Lopes Dos Santos, A., Gourvil, P., Tragin, M., Noël, M.-H., Decelle, J., Romac, S., and Vaulot, D. (2017). Diversity and oceanic distribution of prasinophytes clade VII, the dominant group of green algae in oceanic waters. ISME J. *11*, 512–528.

Pernice, M.C., Forn, I., Gomes, A., Lara, E., Alonso-Sáez, L., Arrieta, J.M., del Carmen Garcia, F.,

Hernando-Morales, V., MacKenzie, R., Mestre, M., et al. (2015). Global abundance of planktonic heterotrophic protists in the deep ocean. ISME J. *9*, 782–792.

Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017). The enigmatic archaeal virosphere. Nat. Rev. Microbiol. *15*, 724–739.

Proud, R., Handegard, N.O., Kloser, R.J., Cox, M.J., and Brierley, A.S. (2019). From siphonophores to deep scattering layers: uncertainty ranges for the estimation of global mesopelagic fish biomass. ICES J. Mar. Sci. *76*, 718–733.

Silsbe, G.M., Behrenfeld, M.J., Halsey, K.H., Milligan, A.J., and Westberry, T.K. (2016). The CAFE model: A net production model for global ocean phytoplankton: NET PHYTOPLANKTON PRODUCTION. Global Biogeochem. Cycles *30*, 1756–1777.

Sissini, M.N., de Barros Barreto, M.B.B., Széchy, M.T.M., de Lucena, M.B., Oliveira, M.C., Gower, J., Liu, G., de Oliveira Bastos, E., Milstein, D., Gusmão, F., et al. (2016). The floating Sargassum (Phaeophyceae) of the South Atlantic Ocean--likely scenarios. Phycologia *56*, 321–328.

Smith, S.V. (1981). Marine macrophytes as a global carbon sink. Science *211*, 838–840.

Turley, C.M., and Mackie, P.J. (1994). Biogeochemical significance of attached and free-living bacteria and the flux of particles in the NE Atlantic Ocean. Mar. Ecol. Prog. Ser. *115*, 191–203.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. Science *348*, 1261605.

Wang, M., and Hu, C. (2017). Predicting Sargassum blooms in the Caribbean Sea from MODIS observations : Sargassum Bloom Prediction. Geophys. Res. Lett. *44*, 3265–3273.

Wang, M., Hu, C., Cannizzaro, J., English, D., Han, X., Naar, D., Lapointe, B., Brewton, R., and Hernandez, F. (2018). Remote Sensing of Sargassum Biomass, Nutrients, and Pigments. Geophys. Res. Lett. *45*, 12,359–12,367.

Whittaker, R.H., and Likens, G.E. (1973). Carbon in the biota. Brookhaven Symp. Biol. 281–302.

Wigington, C.H., Derek, S., Brussaard, C.P.D., Alison, B., Finke, J.F., Fuhrman, J.A., Lennon, J.T., Mathias, M., Suttle, C.A., Charles, S., et al. (2016). Re-examination of the relationship between marine virus and microbial cell abundances. Nature Microbiology *1*, 15024.

Wilson, R.W., Millero, F.J., Taylor, J.R., Walsh, P.J., Christensen, V., Jennings, S., and Grosell, M. (2009). Contribution of Fish to the Marine Inorganic Carbon Cycle. Science *323*, 359–362.