nature geoscience

ARTICLES

https://doi.org/10.1038/s41561-019-0421-8

Important contribution of macroalgae to oceanic carbon sequestration

Alejandra Ortega[1], Nathan R. Geraldi[1], Intikhab Alam[2], Allan A. Kamau[2], Silvia G. Acinas[3], Ramiro Logares[3], Josep M. Gasol[3,4], Ramon Massana[3], Dorte Krause-Jensen[5,6] and Carlos M. Duarte[1,2*]

The role of macroalgae in Blue Carbon assessments has been controversial, partially due to uncertainties about the fate of exported macroalgae. Available evidence suggests that macroalgae are exported to reach the open ocean and the deep sea. Nevertheless, this evidence lacks systematic assessment. Here, we provide robust evidence of macroalgal export beyond coastal habitats. We used metagenomes and metabarcodes from the global expeditions Tara Oceans and Malaspina 2010 Circumnavigation. We discovered macroalgae worldwide at up to 5,000 km from coastal areas. We found 24 orders, most of which belong to the phylum Rhodophyta. The diversity of macroalgae was similar across oceanic regions, although the assemblage composition differed. The South Atlantic Ocean presented the highest macroalgal diversity, whereas the Red Sea was the least diverse region. The abundance of macroalgae sequences attenuated exponentially with depth at a rate of 37.3% km^{-1}, and only 24% of macroalgae available at the surface were expected to reach the seafloor at a depth of 4,000 m. Our findings indicate that macroalgae are exported across the open and the deep ocean, suggesting that macroalgae may be an important source of allochthonous carbon, and their contribution should be considered in Blue Carbon assessments.

Coastal habitats are highly productive ecosystems that contribute greatly to global carbon sequestration[1,2]. Seagrass meadows, salt marshes and mangrove forests have complex root systems that sequester large amounts of carbon in soft sediments within their habitat[3–6]. Macroalgae have been neglected in Blue Carbon assessments[7,8], because most of them lack root systems, grow on rocky substrates, and do not accumulate carbon-rich sediments. However, macroalgae form the most extensive and productive vegetated coastal habitat, exporting over 44% of their primary production[1,7,9]. Calculations suggest that 25% of exported macroalgal carbon is sequestered in long-term reservoirs, such as coastal sediments and the deep sea[1,7].

Based on first-order calculations[7], it is hypothesized that macroalgae globally support an export of 679 TgC yr^{-1}. Most of this carbon is remineralized or grazed in coastal environments, or cast onshore, while 14 TgC yr^{-1} is sequestered in coastal sediments, and 152 TgC yr^{-1} could be sequestered in the deep sea[7]. Although there is a lack of empirical data, these calculations are supported by anecdotal evidence from sightings of long-distance macroalgae rafting[10] and presence in deep-sea sediments[7]. This evidence is dominated by observations of large biomasses of brown macroalgae (Phaeophyta), but observations of red (Rhodophyta) and green (Chlorophyta) macroalgae are few[10]. This evidence imbalance could be related to lineage-specific features of the macroalgae cell wall composition and differences in cell degradation rates[11]. Furthermore, most calculations of macroalgal primary production suggest that macroalgal carbon is exported as dissolved organic carbon and particulate organic carbon (POC)[12,13], which are not visually detectable. An inclusive method, such as the identification of macroalgal environmental DNA (eDNA), could provide evidence of macroalgal carbon export in the ocean, and may allow the required systematic and consistent assessments. eDNA is the DNA left behind by organisms in the surrounding environment, including degraded cell tissues, gametes, animal faeces, and so on. As DNA comprises approximately 3% of cellular organic carbon[14], the presence of macroalgal DNA in waters beyond macroalgal habitats is both an indicator of the presence of the species and evidence (not necessarily quantitative) of the export of macroalgal carbon.

Here, we examined the presence and relative abundances of Rhodophyta, Phaeophyta and Chlorophyta macroalgal eDNA sequences in the ocean. The sequences were derived from hundreds of metagenomes generated by two global expeditions: Tara Oceans[15] and the Malaspina 2010 Circumnavigation[16]. These expeditions surveyed the global ocean from the surface to a depth of 4,000 m, and sequenced the particulate material present in environmental water samples[17,18] (see Methods). Although the expeditions primarily assessed the microbial and planktonic diversity, they also generated a global DNA resource that allows the identification of multicellular eukaryotes. We exploited the potential of this eukaryotic eDNA resource to explore the presence of macroalgae in the global ocean. This holistic approach has not been attempted before, but is semiquantitative and consistent for evaluating the hypothesis that macroalgal material is broadly exported across the global ocean.

We identified macroalgae using two global ocean datasets. The first included 163 metabarcodes of amplicon 18S rDNA from Tara Oceans[19]. The second included 417 metagenomes pooled from the Tara Oceans[20] and Malaspina[21] expeditions (see Methods). We used two different strategies for the second dataset: (1) a query targeting

[1]Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [2]Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [3]Institut de Ciències del Mar, CSIC, Barcelona, Spain. [4]Centre for Marine Ecosystem Research, Edith Cowan University, Western Australia, Australia. [5]Arctic Research Centre, Department of Bioscience, Aarhus University, Aarhus, Denmark. [6]Department of Bioscience, Aarhus University, Silkeborg, Denmark. *e-mail: carlos.duarte@kaust.edu.sa

Table 1 | Relative abundance of macroalgal DNA

Dataset	Lineage	Order	Indian Ocean	Mediterranean Sea	North Atlantic Ocean	North Pacific Ocean	Red Sea	South Atlantic Ocean	Southern Ocean	South Pacific Ocean	Order abundance (%)	Order prevalence (%)
SCG	Chlorophyta	Prasiolales	1.3±0.6	0.9±0.4	2.2±0.5	0.7±0.4	0.3±0.3	1.6±0.3	2.5±1.4	1.2±0.2	11.46	40.05
SCG	Phaeophyta	Ectocarpales	1.1±0.3	4.1±1.4	2.1±0.5	3.7±1.2	0.9±0.3	2.1±0.3	4.3±1.5	2.1±0.4	21.87	49.40
SCG	Phaeophyta	Fragilariales	0.2±0.1	0.4±0.1	0.9±0.3	0.1±0	0.4±0.2	0.6±0.2	-	0.3±0.2	3.11	17.75
SCG	Phaeophyta	Laminariales	0.2±0.1	-	-	0.2±0.2	-	0.2±0.1	-	-	0.77	4.56
SCG	Rhodophyta	Bangiales	-	-	-	-	-	-	0.1±0.1	-	0.14	1.44
SCG	Rhodophyta	Batrachospermales	-	-	-	-	-	0.4±0.3	1.0±0.9	-	1.47	2.88
SCG	Rhodophyta	Bonnemaisoniales	-	-	-	0.1±0.1	-	-	-	-	0.12	0.72
SCG	Rhodophyta	Ceramiales	-	-	-	-	-	-	-	-	0.05	0.96
SCG	Rhodophyta	Corallinales	0.1±0	0.1±0.1	0.3±0.2	0.1±0	-	1.3±0.5	-	0.1±0	2.14	15.83
SCG	Rhodophyta	Cyanidiales	3.0±0.7	1.8±0.7	3.6±0.8	2.5±0.5	7.6±4.9	4.8±1.1	2.6±0.7	7.1±1.2	35.31	72.42
SCG	Rhodophyta	Gelidiales	0.1±0	0.2±0.2	-	0.1±0	-	0.5±0.2	0.1±0.1	0.1±0	1.14	6.24
SCG	Rhodophyta	Gigartinales	1.2±0.2	0.1±0.1	0.8±0.2	1.1±0.3	0.6±0.3	0.7±0.2	0.9±0.6	1.1±0.2	6.90	46.52
SCG	Rhodophyta	Halymeniales	0.1±0.1	-	0.2±0.1	0.4±0.2	-	0.3±0.1	0.2±0.2	0.2±0.1	1.58	9.35
SCG	Rhodophyta	Nemaliales	0.5±0.2	0.1±0	2.3±0.8	1.1±0.4	0.1±0.1	2.3±0.6	2.4±2.1	3.6±0.7	13.22	32.61
SCG	Rhodophyta	Palmariales	-	-	-	-	-	-	0.1±0.1	-	0.18	1.44
SCG	Rhodophyta	Plocamiales	-	-	-	-	-	-	-	-	0.01	0.24
SCG	Rhodophyta	Rhodymeniales	-	-	-	-	-	0.3±0.1	-	0.2±0	0.54	13.19
All genes	Chlorophyta	Prasiolales	0±0.1	-	0±0.1	-	-	0.1±0.3	0.2±0.2	-	1.21	7.91
All genes	Chlorophyta	Ulvales	-	-	-	-	-	0.1±0.1	-	-	0.28	1.44
All genes	Phaeophyta	Ectocarpales	0.4±1.5	0.1±0.4	0.7±2.5	0.6±1.8	0.7±2.0	0.9±2.0	0.3±0.5	0.7±3.1	25.10	30.22
All genes	Phaeophyta	Fucales	-	-	-	-	-	-	-	-	0.06	0.48
All genes	Phaeophyta	Laminariales	-	-	-	-	-	-	-	-	0.07	6.71
All genes	Rhodophyta	Balliales	0.4±0.6	-	-	0.1±0.1	-	0.6±0.8	0.2±0.3	0.4±0.6	4.33	6.71
All genes	Rhodophyta	Bangiales	0.8±1.5	0.3±0.5	0.1±0.5	0.3±0.8	0.1±0.1	0.3±1.0	0.9±1.4	0.2±0.7	11.87	33.09
All genes	Rhodophyta	Ceramiales	-	-	-	-	-	-	0.1±0.1	-	0.15	0.48
All genes	Rhodophyta	Corallinales	-	-	-	-	-	-	-	-	0.04	0.48
All genes	Rhodophyta	Cyanidiales	0.2±0.6	8.3±9.0	0.4±1.6	0.4±1.4	0.4±0.7	0.5±1.5	0.6±1.8	0.2±1.3	32.41	44.12
All genes	Rhodophyta	Gigartinales	0.1±0.2	-	0.1±0.1	0.2±0.2	-	0.1±0.2	0.1±0.1	0.1±0.1	1.72	5.28
All genes	Rhodophyta	Gracilariales	0.1±0.1	-	-	0.3±0.3	-	0.1±0.2	-	-	1.05	1.92
All genes	Rhodophyta	Hapalidiales	-	-	-	-	-	-	-	-	0.08	0.48
All genes	Rhodophyta	Nemaliales	0.1±0.2	0.1±0.2	0.3±0.9	0.1±0.3	0.6±1.3	0.5±1.2	-	0.1±0.3	7.94	24.22
All genes	Rhodophyta	Palmariales	-	-	-	0.2±0.2	2.4±4.6	-	-	-	8.93	3.84
All genes	Rhodophyta	Porphyridiales	-	-	-	-	-	-	-	-	0.04	0.24
All genes	Rhodophyta	Stylonematales	0±0.1	0.2±0.3	0.2±0.5	0.1±0.2	0.1±0.2	0.3±1.0	-	0.1±0.4	4.71	19.42
18S	Chlorophyta	Prasiolales	9.7±2.8	1.0±0.6	0.8±0.5	-	1.5±0.5	13.5±7.2	10.7±3.5	4.2±1.1	53.20	66.23
18S	Rhodophyta	Ceramiales	-	16.8±14.6	-	-	-	-	-	-	21.63	12.99
18S	Rhodophyta	Gigartinales	-	0.2±0.2	-	-	-	-	-	-	0.29	2.60
18S	Rhodophyta	Porphyridiales	0.4±0.3	8.4±8.0	1.8±3.5	-	-	4.3±3.4	0.5±0.3	0.4±0.1	20.16	35.06
18S	Rhodophyta	Rhodymeniales	-	3.7±3.5	-	-	-	-	-	-	4.73	3.90

Order abundance indicates the percentage of each order among total macroalgal DNA sequences per dataset. Order prevalence indicates the percentage of metagenomes/metabarcodes (*n*) where the order was present. The number (*n*) of metagenomes (for both SCG/all genes) or metabarcodes (18S) per region, respectively, were: 87 and 11 for the Indian Ocean; 19 and 18 for the Mediterranean Sea; 44 and 2 for the North Atlantic Ocean; 101 and 21 for the South Atlantic Ocean; 51 and 0 for the North Pacific Ocean; 88 and 18 for the South Pacific Ocean; 13 and 2 for the Red Sea; and 14 and 6 for the Southern Ocean. Cyanidiales represented 35% of the SCG dataset (reads per million ± s.e.) and was present in 72% of the SCG metagenomes. Prasiolales dominated the 18S dataset (53% of the total sequences, metagenomic Illumina tags ± s.e.) and was found in 66% of the metabarcodes.

all genes; and (2) restriction of the query to the top-four single-copy protein-encoding genes (throughout this work SCG refers specifically to these four) available in the gene catalogue of both expeditions. Since macroalgae taxonomy is not well covered in barcoding and genome reference libraries[22], we used order instead of species as the taxonomic level for macroalgae identification.

Tracing macroalgae

Combining the results of the three independent datasets, 24 macroalgae orders were identified within the particulate organic matter (POM) of the water column. Both metagenomic approaches (all genes and SCG) delivered 17 orders, of which ten were shared among the two approaches (Table 1). Only six orders were detected

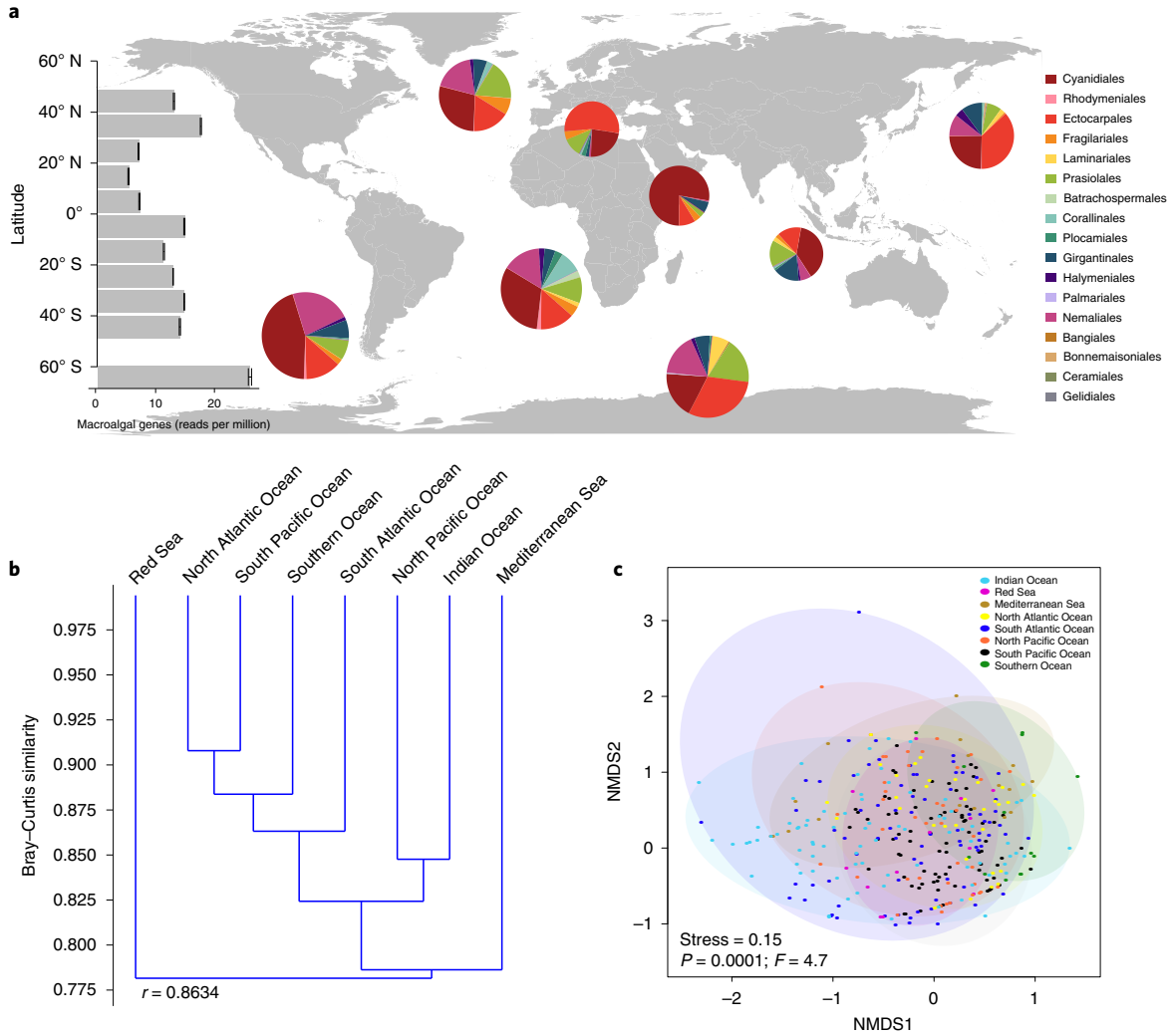

Fig. 1 | Assemblage of macroalgae in the ocean. a, Global distribution of macroalgae. Pie sizes represent DNA abundance per region, with highest abundance in the South Pacific Ocean (17%) and lowest abundance in the Mediterranean Sea and Indian Ocean (8% each). The bar graph shows the latitudinal distribution (means ± s.e.) of total macroalgal DNA, with 50% abundance beyond 40° N and 40° S. **b**, Bray–Curtis cluster showing similarity > 78% among all regions (cophenetic correlation, $r = 0.8634$) . **c**, nMDS comparing macroalgal assemblage across oceanic regions, each one indicated by a shaded colour.

in the amplicon 18S rDNA metabarcodes, all of which were found in the metagenomes. Rhodophyta was the most common macroalgal lineage (18 orders: 12 in all genes, 13 in SCG and four in the 18S dataset), followed by Phaeophyta (four orders: two unique in each metagenomic approach) and Chlorophyta (two orders: two in all genes, and one shared in both SCG and 18S datasets; Table 1).

The relative abundance of macroalgal DNA varied between oceanic basins and datasets. The Mediterranean Sea presented the highest abundance of sequences in both the 18S and all genes datasets, while the South Atlantic Ocean was the most abundant in the SCG dataset. The basins with the fewest sequences were the Red Sea in the 18S dataset, the Southern Ocean for all genes and the Mediterranean Sea in the SCG dataset. Similarly, the relative abundance of sequences per order differed greatly. Cyanidiales (Rhodophyta) and Ectocarpales (Phaeophyta) jointly accounted for 57% of the macroalgal sequences in both metagenomic datasets, although they were absent from the amplicon 18S dataset, whose most abundant order was Prasiolales (Chlorophyta), with 53% of all macroalgal sequences (Table 1).

Our pioneering attempt to trace macroalgal eDNA from POM in the global ocean is challenging for two reasons. First, the phylogenetic diversity of macroalgae is so great that the three lineages are as distant from each other as mushrooms are from elephants[8]. Second, macroalgal sequences are poorly represented in reference libraries. Metagenomic and metabarcoding identification is restricted to previously sequenced taxa that are available in published databases. Sequencing efforts on macroalgae are rather limited, with only one full genome sequenced[23]. Half of the 24 orders identified here are not included in the SILVA 18S rDNA reference library (http://www.arb-silva.de). Furthermore, SILVA includes only 1,068 macroalgae species, compared with 12,471 species reported in AlgaeBase and 27,500 described species[22]. SILVA under-represents green and brown macroalgae compared with red algae: Chlorophyta and Phaeophyceae have 46 and 84 entries for macroalgae, respectively, while Rhodophyta has 938 entries (searched in July 2018). Analogously, macroalgae do not have any single-copy protein-encoding gene reported in the EggNOG database (http://eggnogdb.embl.de), as most proteins are reported for model organisms such as *Oryza sativa*, *Arabidopsis thaliana* or *Saccharomyces cerevisiae*. Because of this scarcity in macroalgae reference sequences, there is an underestimation of macroalgae (false negatives) and a bias in the taxonomic representation of the macroalgae contributing

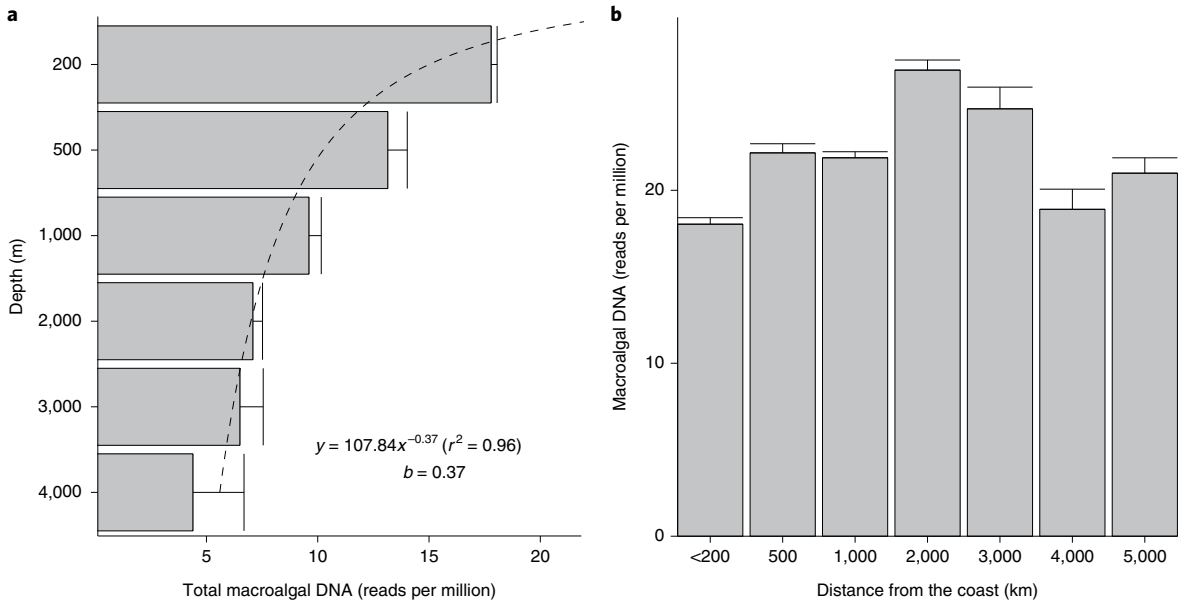

Fig. 2 | Export of macroalgae to the deep and open ocean. a, Vertical attenuation profile of macroalgal DNA in the Malaspina SCG dataset. The dashed line shows the fitting curve based on the power law equation of *y*. The calculated attenuation coefficient is indicated by *b* (see Methods). **b**, Horizontal export of macroalgal DNA from the shoreline (continent or island) to the open ocean (sampling point), based on the Tara and Malaspina epipelagic metagenomes (0–200 m depth). Data are means ± s.e.

to the POM. There is a need for enhanced molecular resources for macroalgae, especially for single markers. A single marker (that is, 18S rDNA gene) enhances accurate identification to the species level, and could draw phylogenetic relationships among lineages. A robust genomic reference will allow the detection of species in the POC and dissolved organic carbon pools, enabling the use of eDNA-based approaches to assign relative contributions of species to the carbon available in the ocean.

Macroalgae taxonomic identification in all datasets was performed by matching the sequences against available DNA references. Macroalgae sequences were less abundant in the 18S metabarcoding dataset, with only 29% of orders available in the metagenomes. Thus, the metagenomes make it easier to find macroalgal DNA in the water column, given the poor and highly unbalanced representation of macroalgae in the SILVA 18S library.

Macroalgal material is likely to be exported from their coastal habitats as whole thalli or fragments that either degrade progressively or are rapidly delivery to the deep sea[7]. Although marine eDNA decays within a few days[24,25], the drifting macroalgal biomass[7] is constantly leaving traces of its DNA. eDNA recovered from metagenomes is the snapshot evidence of the macroalgal biomass exported to the sampling location from the coastal habitat. However, it is uncertain whether the relative abundance of sequences per order truly reflects the contribution of each order within the macroalgal export flux. The focus on metagenomic SCGs provides a parsimonious approach for assessing the relative abundance of macroalgae. A single-copy gene occurs once in the genome, accounts for a single cell and represents one individual in microbial communities[26]. In multicellular organisms, the relative abundance of DNA sequences from SCGs may be scaled to the relative number of cells (and amount of biomass) available per taxon. Thus, the abundance pattern of macroalgal SCGs from different taxa may be expected to correlate with their contribution to carbon export.

Given these caveats for metagenomes, and considering that the 18S metabarcodes were limited to fewer samples, we chose the SCG dataset for further analysis of macroalgal order diversity and macroalgal biomass export in the open and deep ocean. We believe that the SCG approach is probably less biased and more informative than the other two approaches.

Macroalgal diversity in the ocean

Macroalgal taxonomic composition in the SCG dataset was similar across oceanic regions. Cyanidiales and Ectocarpales were the most ubiquitous and abundant orders across all of the basins. Cyanidiales represented 35% of macroalgal DNA sequences. This result was unexpected but may possibly be related to the fact that Cyanidiales is the earliest Rhodophyta and other orders could share enough nucleotides in the sequences that may be identified as Cyanidiales. Nevertheless, we aligned these DNA sequences, and the phylogeny separates Rhodophyta orders (Supplementary Fig. 1). Furthermore, Cyanidiales is known for its metabolic capacities and ability to colonize extreme habitats[27]. Ectocarpales—the most diverse order of Phaeophyta (774 species in AlgaeBase[28])—accounted for 22% of the DNA sequences (Table 1 and Fig. 1a). The Atlantic and North Pacific Oceans were the most diverse regions, while the Red Sea (the smallest basin sampled) was the least diverse (Supplementary Table 1 and Fig. 1a). The South Atlantic Ocean displayed the highest percentage of macroalgal DNA (17% of the total across all basins), while the lowest was found in the Mediterranean Sea and the Indian Ocean (8% each). A high abundance of macroalgae was observed poleward of 40° in both the Northern (21%) and Southern (28%) Hemisphere (Fig. 1a), possibly reflecting high local production of macroalgae at these latitudes. The Arctic supports abundant macroalgae populations along its extensive rocky coastline[29], and the Norwegian Atlantic current may collect significant inputs of boreal macroalgal detritus. Similarly, there is evidence of the export of Antarctic kelps—brown macroalgae of the order Laminariales—that could potentially be transported over long distances by the Antarctic Circumpolar Current[30]. In addition, macroalgal material may be preserved longer at low water temperatures than at the warmer temperatures found at tropical latitudes[31]. Since many species contain air vesicles that confer buoyancy, polar latitudes could be a dead end for macroalgal material, as has been shown to be the case for plastic accumulation driven by surface circulation[32].

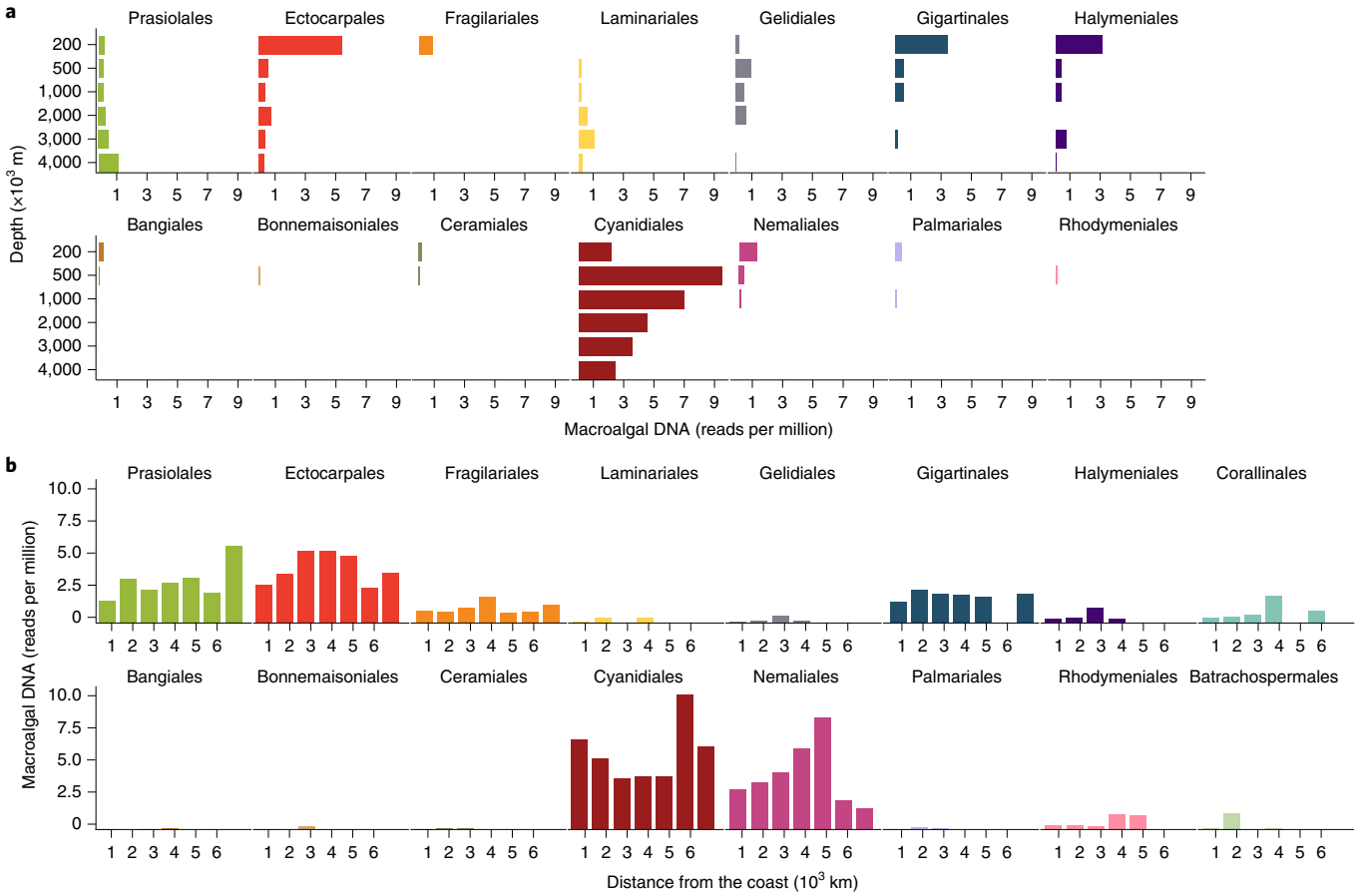

Fig. 3 | Oceanic export of macroalgal DNA relative abundance per order. a, Depth attenuation of macroalgae from the surface to the bathypelagic zone (4,000 m), using Malaspina metagenomes. Most orders attenuated with depth, with the exception of Prasiolales and Laminariales. **b**, Macroalgal DNA export from the shoreline to the open ocean, based on the Tara and Malaspina epipelagic metagenomes (0–200 m depth). The presence of macroalgae is ubiquitous in the open ocean.

One-way permutational multivariate analysis of variance (PERMANOVA) revealed significant differences in the SCG macroalgal DNA assemblage across oceans ($P=0.0001$; degrees of freedom (d.f.) = 7,347; pseudo-F statistic (F) = 4.7). The Red Sea, Indian Ocean and South Pacific Ocean were significantly different from the other oceanic regions (pairwise $P<0.005$; Supplementary Table 2). Nevertheless, cluster analysis and non-metric multi-dimensional scaling (nMDS) ordination indicated similarities in the assemblages (Fig. 1b,c). Most regions were above 82% similarity, with the Red Sea and Mediterranean Sea (77% similarity) as exceptions. Differences between overall PERMANOVA and ordination indicate a dispersion effect, as confirmed by the significant difference in variance between the groups (analysis of homogeneity of multivariate dispersion (PERMDISP), $P<0.0001$; d.f. = 7,347; $F=5.7$).

Export of macroalgae throughout the water column

The Malaspina expedition sampled eDNA from the surface to 4,000 m, while for Tara the maximum sampling depth was 1,000 m. Consequently, analyses of oceanic macroalgal abundance include only the Malaspina dataset. The order diversity of macroalgae varied between depths (PERMANOVA: $P=0.001$; d.f. = 2,352; $F=43.6$; pairwise $P<0.05$; PERMDISP: $P<0.0001$; d.f. = 2,352; $F=31.7$). The epipelagic zone (0–200 m) was the most diverse, while the least diverse was the mesopelagic zone (200–1,000 m; Supplementary Table 3). The relative abundance of macroalgal DNA (and probably macroalgal carbon) attenuates exponentially with depth at a rate of 37.3% km^{-1} (Fig. 2a). This value is much lower than the attenuation rate of sinking POC flux in the Northeast Pacific Ocean down to 5,000 m (86% km^{-1}, based on data from Martin et al.[33]). However, a lower value for the global ocean attenuation rate is fairly expected due to the refractory nature of macroalgae carbon, which degrades slower compared with planktonic POC[34]. These results provide large-scale quantitative evidence of macroalgal transport to the deep sea, validating previous assumptions of vertical export[7].

Most macroalgae grow in coastal areas. Exceptions are the drifting Sargasso Sea, and macroalgae living on shallow oceanic seamounts[35,36]. Oceanic and biological processes (for example, storms and senescence) promote coastal detachment, dispersion and export of macroalgae to the open ocean[7,37]. In contrast with the exponential attenuation by depth, there was no difference in macroalgal abundance from the shoreline to distances up to 4,860 km (PERMANOVA: $P=0.194$; d.f. = 6,223; $F=1.2$; Fig. 2b). This observation corroborates the estimated widespread export of macroalgal material to the open ocean, hitherto based on anecdotal evidence.

Rhodophyta can tolerate long periods of darkness and remain photosynthetic at great depths[38,39]. Furthermore, these macroalgae cover the largest geographical extent and support the largest global production[38]. Thus, Rhodophyta would be estimated to export more material than Phaeophyta and Chlorophyta. Our results confirm these assertions: several red algae were present at high depths, and 63% of the DNA sequences belonged to Rhodophyta, compared with 26% for Phaeophyta and 11% for Chlorophyta (Table 1 and Fig. 3). Likewise, Rhodophyta were taxonomically more abundant than Phaeophyta and Chlorophyta (13 > 3 > 1 order, respectively).

This richness is expected: AlgaeBase shows a greater diversity of red algae (6,245 classes; 30 orders) than brown (1,792 classes; 13 orders) or green algae (546 classes; 15 orders)[8,22]. The low oceanic richness of Chlorophyta in the exported POM could be related to morphological and biochemical features. Rhodophyta and Phaeophyta contain taxon-specific polysaccharides that provide structural complexity and recalcitrancy[40]. Fucoidans (in brown algae) and carrageenans (in red algae) bind to the cell wall and protect it from desiccation and cell invasion by microbes, hence delaying degradation[11,41–43]. These features are absent in green algae[41]. Such recalcitrance-promoting compounds may enhance the long-distance transport of Rhodophyta and Phaeophyta, as supported by their prevalence in the oceanic POM pool.

Implications for Blue Carbon assessment

Our findings show the ubiquitous presence of macroalgal DNA in the ocean up to a depth of 4,000 m, and 4,860 km away from the nearest coastline. The attenuation rate of macroalgae (37.3% km^{-1}) implies that 69% of the macroalgal DNA available at the surface will sink below 1,000 m. Oceanic models show that the carbon reaching a depth of 1,500 m is sequestered close to permanent timescales[44] in terms of climate change mitigation. Hence, the macroalgal material (and organic carbon) that reaches 1,000 m (the boundary between the mesopelagic and bathypelagic layers[45]) will be sequestered and prevented from exchanging with the atmosphere over extended timescales[7,44]. Moreover, 24% of macroalgal DNA sequences sinking from the surface will be expected to reach the seafloor (assuming a mean oceanic depth of 3,800 m). Our results also reveal an increase in the relative abundance of Laminariales (for instance, kelp) DNA in POM between 3,000 and 4,000 m (Fig. 3a), consistent with the reported bedload bulk transport of kelp to the deep sea[7]. This transport is influenced by episodic storm-driven events[7,46] that detach and rapidly sink macroalgae; this rapid sink is due to the presence of heavy rocky substrate retained by macroalgae in their holdfast. Submarine canyons support intense bedload fluxes of kelp, thereby delivering macroalgae (along with their DNA sequences and carbon) directly into the deep sea[7,47–49]. Through this mechanism, a larger biomass of Laminariales is delivered to the deep sea, while the remaining orders progressively degrade into smaller and smaller fragments, thus attenuating exponentially with increasing depth.

While the global ocean metagenomes analysed here were produced to explore the oceanic microbiome, the data also allow the detection of eukaryotic organisms such as macroalgae. Metagenomes are an unexplored tool for fingerprinting the contributions of different organisms to POM in the ocean. This research is a first step supporting the role of macroalgae as an important allochthonous source of Blue Carbon sequestered in the deep sea. Our eDNA approach provides robust evidence of the widespread oceanic presence of macroalgae, and our data support the hypothesis of macroalgae export to the open and deep ocean hitherto based on estimations[1,7]. As DNA is also cellular organic carbon[14], we infer that the presence of the taxa evidences the export of macroalgal carbon. Nevertheless, calculations of the macroalgal carbon exported to the ocean require experimentally determined ratios between carbon and DNA content per taxon, which are currently unknown. Although the ultimate fate of oceanic macroalgal material remains uncertain, it is clear that a significant fraction reaches oceanic sinks while another is grazed and degraded by bacteria, thus subsidizing oceanic food webs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41561-019-0421-8.

Received: 10 December 2018; Accepted: 28 June 2019;
Published online: 05 August 2019

References

1. Duarte, C. M. & Cebrián, J. The fate of marine autotrophic production. *Limnol. Oceanogr.* **41**, 1758–1766 (1996).
2. Duarte, C. M. & Krause-Jensen, D. Export from seagrass meadows contributes to marine carbon sequestration. *Front. Mar. Sci.* **4**, 13 (2017).
3. McLeod, E. et al. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO_2. *Front. Ecol. Environ.* **9**, 552–560 (2011).
4. Fourqurean, J. W. et al. Seagrass ecosystems as a globally significant carbon stock. *Nat. Geosci.* **5**, 505–509 (2012).
5. Duarte, C. M., Kennedy, H., Marbà, N. & Hendriks, I. Assessing the capacity of seagrass meadows for carbon burial: current limitations and future strategies. *Ocean Coast. Manage.* **83**, 32–38 (2013).
6. Donato, D. C. et al. Mangroves among the most carbon-rich forests in the tropics. *Nat. Geosci.* **4**, 293–297 (2011).
7. Krause-Jensen, D. & Duarte, C. M. Substantial role of macroalgae in marine carbon sequestration. *Nat. Geosci.* **9**, 737–742 (2016).
8. Krause-Jensen, D. et al. Sequestration of macroalgal carbon: the elephant in the Blue Carbon room. *Biol. Lett.* **14**, 20180236 (2018).
9. Duarte, C. M. Reviews and syntheses: hidden forests, the role of vegetated coastal habitats in the ocean carbon budget. *Biogeosciences* **14**, 301–310 (2017).
10. Garden, C. J. & Smith, A. M. Voyages of seaweeds: the role of macroalgae in sediment transport. *Sediment. Geol.* **318**, 1–9 (2015).
11. Kloareg, B. & Quatrano, R. S. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanogr. Mar. Biol.* **26**, 259–315 (1988).
12. Barrón, C., Apostolaki, E. T. & Duarte, C. M. Dissolved organic carbon fluxes by seagrass meadows and macroalgal beds. *Front. Mar. Sci.* **1**, 42 (2014).
13. Krumhansl, K. A. & Scheibling, R. E. Production and fate of kelp detritus. *Mar. Ecol. Prog. Ser.* **467**, 281–302 (2012).
14. Landenmark, H. K. E., Forgan, D. H. & Cockell, C. S. An estimate of the total DNA in the biosphere. *PLoS Biol.* **13**, e1002168 (2015).
15. Karsenti, E. et al. A holistic approach to marine eco-systems biology. *PLoS Biol.* **9**, e1001177 (2011).
16. Duarte, C. M. Seafaring in the 21st century: the Malaspina 2010 Circumnavigation Expedition. *Limnol. Oceanogr.* **24**, 11–14 (2015).
17. Salazar, G. et al. Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J.* **10**, 596–608 (2016).
18. Hingamp, P. et al. Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* **7**, 1678–1695 (2013).
19. De Vargas, C. et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).
20. Pesant, S. et al. Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* **2**, 150023 (2015).
21. Sánchez, P. et al. Dataset: Common photosynthetic enzymes from 174 metagenomes from the Malaspina Expedition 2010. Supplement to: Ortega et al. (2019). Important contribution of macroalgae to oceanic carbon sequestration. *Zenodo* https://doi.org/10.5281/zenodo.2596829 (2019).
22. Guiry, M. D. How many species of algae are there? *J. Phycol.* **48**, 1057–1063 (2012).
23. Cock, J. M. et al. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621 (2010).
24. Collins, R. A. et al. Persistence of environmental DNA in marine systems. *Commun. Biol.* **1**, 185 (2018).
25. Thomsen, P. F. et al. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE* **7**, e41732 (2012).
26. Roux, S., Enault, F., le Bronner, G. & Debroas, D. Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. *FEMS Microbiol. Ecol.* **78**, 617–628 (2011).
27. Seckbach, J. & Chapman, D. J. *Red Algae in the Genomic Age* (Springer, 2010).
28. Guiry, M. D. & Guiry, G. M. *AlgaeBase* (National Univ. Ireland, 2013); http://www.algaebase.org
29. Krause-Jensen, D. & Duarte, C. M. Expansion of vegetated coastal ecosystems in the future Arctic. *Front. Mar. Sci.* **1**, 77 (2014).
30. Kaehler, S., Pakhomov, E. A., Kalin, R. M. & Davis, S. Trophic importance of kelp-derived suspended particulate matter in a through-flow sub-Antarctic system. *Mar. Ecol. Prog. Ser.* **316**, 17–22 (2006).
31. Kelaher, B. P., Coleman, M. A. & Bishop, M. J. Ocean warming, but not acidification, accelerates seagrass decomposition under near-future climate scenarios. *Mar. Ecol. Prog. Ser.* **605**, 103–110 (2018).
32. Cózar, A. et al. The Arctic Ocean as a dead end for floating plastics in the North Atlantic branch of the Thermohaline Circulation. *Sci. Adv.* **3**, e1600582 (2017).

33. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. VERTEX: carbon cycling in the northeast Pacific. *Deep Sea Res. Pt II* **34**, 267–285 (1987).
34. Enríquez, S., Duarte, C. M. & Sand-Jensen, K. A. J. Patterns in decomposition rates among photosynthetic organisms: the importance of detritus C:N:P content. *Oecologia* **94**, 457–471 (1993).
35. Carpenter, E. J. & Cox, J. L. Production of pelagic *Sargassum* and a blue-green epiphyte in the western Sargasso Sea. *Limnol. Oceanogr.* **19**, 429–436 (1974).
36. Woodborne, M. W., Rogers, J. & Jarman, N. The geological significance of kelp-rafted rock along the west coast of South Africa. *Geo-Mar. Lett.* **9**, 109–118 (1989).
37. Garden, C. J., Currie, K., Fraser, C. I. & Waters, J. M. Rafting dispersal constrained by an oceanographic boundary. *Mar. Ecol. Prog. Ser.* **501**, 297–302 (2014).
38. Gattuso, J. P. et al. Light availability in the coastal ocean: impact on the distribution of benthic photosynthetic organisms and contribution to primary production. *Biogeosciences* **3**, 895–959 (2006).
39. Littler, M. M., Littler, D. S., Blair, S. M. & Norris, J. N. Deepest known plant life discovered on an uncharted seamount. *Science* **227**, 57–59 (1985).
40. Trevathan-Tackett, S. M. et al. Comparison of marine macrophytes for their contributions to blue carbon sequestration. *Ecology* **96**, 3043–3057 (2015).
41. Percival, E. The polysaccharides of green, red and brown seaweeds: their basic structure, biosynthesis and function. *Br. Phycol.* **14**, 103–117 (1979).
42. Shukla, P. S., Borza, T., Critchley, A. T. & Prithiviraj, B. Carrageenans from red seaweeds as promoters of growth and elicitors of defense response in plants. *Front. Mar. Sci.* **3**, 81 (2016).
43. Berteau, O. & Mulloy, B. Sulfated fucans, fresh perspectives: structures, functions, and biological properties of sulfated fucans and an overview of enzymes active toward this class of polysaccharide. *Glycobiology* **13**, 29R–40R (2003).
44. Herzog, H., Caldeira, K. & Reilly, J. An issue of permanence: assessing the effectiveness of temporary carbon storage. *Clim. Change* **59**, 293–310 (2003).
45. Robinson, C. et al. Mesopelagic zone ecology and biogeochemistry—a synthesis. *Deep Sea Res. Pt II* **57**, 1504–1518 (2010).
46. Dierssen, H. M., Zimmerman, R. C., Drake, L. A. & Burdige, D. J. Potential export of unattached benthic macroalgae to the deep sea through wind-driven Langmuir circulation. *Geophys. Res. Lett.* **36**, L04602 (2009).
47. De Leo, F. C., Smith, C. R., Rowden, A. A., Bowden, D. A. & Clark, M. R. Submarine canyons: hotspots of benthic biomass and productivity in the deep sea. *Proc. R. Soc. Lond. B* **277**, 2783–2792 (2010).
48. Canals, M. et al. Flushing submarine canyons. *Nature* **444**, 354–357 (2006).
49. Harrold, C. & Lisin, S. Radio-tracking rafts of giant kelp: local production and regional transport. *J. Exp. Mar. Biol. Ecol.* **130**, 237–251 (1989).

Acknowledgements

We thank the Tara Oceans Consortium for data availability. This research was supported by the Malaspina 2010 expedition, funded by the Spanish Ministry of Economy and Competitiveness through the Consolider-Ingenio programme to C.M.D. (reference: CSD2008-00077); CARMA, funded by the Independent Research Fund Denmark to D.K.-J. (reference: 8021-00222B); and King Abdullah University of Science and Technology's project BAS/1/1071-01-01 to C.M.D. We thank all of the scientists and crew for support during sample collection on the Malaspina 2010 cruise, and especially E. Borrull, C. Díez-Vives, E. Lara, D. Vaqué, G. Salazar and F. Cornejo-Castillo for DNA sampling. The authors are grateful to the KAUST Supercomputing Laboratory (KSL) for the resources provided.

Author contributions

C.M.D. and D.K.-J. conceived the research. J.M.G., S.G.A., R.L., R.M., I.A. and A.A.K. produced and curated the data. A.O., C.M.D., N.R.G. and I.A. conducted the data analysis. A.O. and C.M.D. wrote the manuscript. All co-authors contributed to improving the manuscript and approved the submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41561-019-0421-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.M.D.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

Macroalgae taxa are included in two of the eight major lineages of the Eukaryota domain[50], where they belong to four kingdoms, 15 phyla and 54 classes[22]. Marine macroalgae are found in three phyla (Rhodophyta, Phaeophyta and Chlorophyta), which also contain microalgae. Chlorophyta (green algae) are closer to vascular plants than to Rhodophyta (red algae) or Phaeophyta (brown algae), which are closer to moulds than to other macroalgae[22,50,51]. Macroalgae groups have broad differences in cell wall composition[11]. Even the same order of algae can have strong divergences among genera[52]. Thus, macroalgae classification is very diverse, and the term 'macroalgae' describes functional groups that are not necessarily related phylogenetically to each other—even at phylum level. Identification of macroalgal DNA sequences is a challenging process: there is no universal gene marker[53], and barcoding attempts are limited to certain groups[53,54]. 18S rDNA barcoding resources are poorly represented: only 3.8% of macroalgae are reported in the SILVA database (1,068 of 27,500 described species[22]; http://www.arb-silva.de (searched in July 2018)). Nevertheless, the available (and limited) molecular resources based on a single gene marker are an important tool for accurately identifying macroalgae, and also for drawing phylogenetic conclusions about these taxa.

However, less strict approaches can also be used to identify marine macroalgae groups when resolution at the species level is not required. Since DNA represents 3% of cellular organic carbon[14], here we infer the carbon export of marine macroalgae phyla with the presence of macroalgal DNA in the water column. We investigated the occurrence of macroalgae in the open and deep ocean using global metagenomes and metabarcodes generated by the Tara Oceans[15] and Malaspina 2010 Circumnavigation[16] expeditions.

Sample description. Tara sampling covered epipelagic and mesopelagic zones (5–1,000 m) across eight oceanic regions (the North Atlantic Ocean, South Atlantic Ocean, North Pacific Ocean, South Pacific Ocean, Indian Ocean, Southern Ocean, Mediterranean Sea and Red Sea), with a total of 210 sampling stations. Each location was sampled at different depths (surface water layer, 3–7 m; deep chlorophyll maximum layer, 30–70 m; mesopelagic zone, 400–1,000 m) using CTD and Niskin bottle Rosette sampling system[17,18,20]. We used 243 metagenomic samples that targeted the gene pool of viral to metazoan plankton, using multiple filters to isolate distinct size-fractions of the suspended particle pool (0.1–0.22 µm 20 samples, <0.22 µm 45 samples, 0.22–0.45 µm 18 samples, 0.45–0.8 µm 21 samples, 0.22–1.6 µm 36 samples, and 0.22–3 µm 103 samples)[20]. We used 163 metabarcodes from the 18S rDNA aimed at piconano- to meso-plankton communities (size fractions: <0.8 µm, 28 samples; 0.22–3 µm, one sample; 0.8–5 µm, 60 samples; 0.8–20 µm, six samples; 5–20 µm, 23 samples; 20–180 µm, 31 samples; 180–2,000 µm, 14 samples)[19]. Water samples were kept at −20 °C on board and at −80 °C in the laboratory until DNA extraction, then DNA was kept at −20 °C until sequencing[55]. Detailed sampling procedures and methods are available for metagenomes[20] and for 18S rDNA amplicons[19].

The Malaspina expedition sampled open-ocean waters from the surface to a depth of 4,018 m, with emphasis on the bathypelagic zone (1,000-4,000 m)[16]. Water samples were collected using CTD and Niskin bottle Rosette sampling systems, at 70 sampling stations across the oceans, and grouped into the eight oceanic regions used by Tara (see above). Filters containing the particle pools sampled in the water were flash frozen in liquid nitrogen and stored at −80 °C until DNA extraction and further sequencing[56]. We used 174 metagenomic[21] samples that targeted free-living bacteria to picoeukaryotes and nanoeukaryotes (size fractions: 0.2–0.8 µm, 29 samples; 0.2–3.0 µm, 100 samples; 0.8–20 µm, 31 samples; 3–20 µm, 14 samples).

Together, these expeditions used massive DNA sequencing and generated hundreds of metagenomes for the assessment of oceanic microbial and planktonic diversity[18]. Tara also generated a global eukaryotic DNA resource based on 18S-V9 rDNA amplicons[11]. This data collection did not aim to survey macro-organisms. Nevertheless, we exploited their potential to reveal macroalgal genes. Data came from 153 Tara and 65 Malaspina sampling stations across the oceans, from the surface to a depth of 4,000 m (Supplementary Fig. 2).

We identified DNA sequences of Rhodophyta, Phaeophyta and Chlorophyta using two datasets: (1) amplicon 18S rDNA-based metabarcodes from Tara Oceans; and (2) metagenomes that contain the whole gene pool from both Tara Oceans and Malaspina. These macroalgal DNA sequences belonged to 20,212 unique genes available in the reference gene catalogues of both expeditions (Supplementary Table 4). We believe that this holistic approach has not been tried before.

Amplicon 18S data extraction. For the first dataset (denoted 18S), single amplicon reads were extracted from 163 Tara[57] metabarcodes of the 18S rDNA V9 hypervariable loop. Metabarcodes were blasted against the SILVA 18S rDNA database (SILVA release 132; http://www.arb-silva.de). Macroalgae taxonomy is not well described[22], and sequences from the 18S rDNA are scarce in the SILVA database (only 3.8% representation; searched in July 2018); thus, to avoid false negative results, we chose order rather than species as the taxonomic level. The search was taxonomically restricted to the taxa Viridiplantae, Stramenopiles and Rhodophyta. The resulting taxonomic list was filtered manually by choosing all macroalgae orders whose sequences presented an identity percentage cut-off of >90%. A cut-off of 90% is above the accepted threshold for order level (84–90%[58–62]).

Malaspina 18S rDNA metabarcodes, though available[63], were excluded for several reasons: Malaspina sequenced the 18S rDNA V4 region, and the sampling and sequencing effort was much lower than in Tara. The contribution of Malaspina to the amplicon 18S dataset was limited to only seven samples presenting any macroalgae sequence, in contrast with 78 samples from Tara.

Metagenomic data extraction. For the second dataset (denoted as metagenomes), we used 243 Tara[20] and 174 Malaspina[21] metagenomes to find macroalgal DNA sequences in the open ocean. We used two different strategies: (1) targeting all genes; and (2) restricting the query to the top-four single-copy protein-encoding genes (SCG) available in the gene catalogues of both expeditions. Each strategy generated its own new dataset. Metagenomic data were analysed using the Dragon Metagenomics Analysis Platform (DMAP; http://www.cbrc.kaust.edu.sa/dmap). DMAP re-annotated Tara Oceans and Malaspina metagenomic gene catalogues, keeping the original reads based on gene abundance for each sample (units are in reads per million).

DMAP uses the UniProt Knowledgebase as a reference database to compare genes from the Tara and Malaspina gene catalogues. To assign taxonomy and generic functional roles, DMAP uses high-throughput BLASTp, which is examined to traverse lowest common ancestor along the best hits. Specific functional roles are assigned using BLASTp against Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologues from the KEGG database. This taxonomic and functional role information is indexed for all genes, and made available for lookups and sample comparisons in the Compare module of DMAP. In this module, we restricted both metagenomic strategies (all genes and SCGs) to the taxa Viridiplantae (DMAP filter taxID: 33090), Stramenopiles (taxID: 33634) and Rhodophyta (taxID: 2763); the search was restricted to coverage and identity percentage cut-offs ≥90%. A higher cut-off recovers fewer sequences (false negatives). The identification of macroalgae in the metagenome dataset is based on protein similarity. Proteins are very conserved at higher taxonomic levels, thus a threshold of 90% is above the mean percentage identity for proteins (70%)[64].

The SCG strategy included additional steps. Initially, we wanted to restrict the search to SCGs specifically from Chlorophyta, Rhodophyta or Phaeophyta, but there were none available in the reference database (EggNOG[65]; searched in February 2018). Thus, we used the KEGG orthologue module of DMAP to search for the top-four SCGs present in Viridiplantae, Stramenopiles and Rhodophyta within the expeditions' gene catalogues. Back to the Compare module, we individually restricted the SCG search to each of the following top-four protein-encoding genes: NADH:ubiquinone reductase (EC: 1.6.5.3), *N*-acetyl-gamma-glutamyl-phosphate reductase (EC: 1.2.1.38), DNA-directed RNA polymerase (EC: 2.7.7.6) and non-specific serine/threonine protein kinase (EC: 2.7.11.1).

Order was used as the level for taxonomic assignment because most macroalgae species have an incomplete genome reference library and undescribed taxonomy[22]. The initial search using species as the taxonomic level returned false negative BLAST hits. For instance, when the search was restricted to a few species of the order Ectocarpales, DMAP did not return any sequence because the dataset is incomplete; sequences were returned when we searched directly for the order. The databases include unknown or uncultured sequences that are assigned to higher taxonomic ranks; for example, order. The search generated a taxonomic list, where we manually filtered all macroalgae orders that returned sequences.

Data analyses. A list of macroalgae orders and the relative abundances of the sequences was obtained from each dataset (18S metabarcodes, metagenomes for all genes and metagenomes for SCGs). Relative abundance is reported as reads per million for the metagenomes, and as metagenomic Illumina tags for the 18S dataset. Each sample included information on depth, size fraction and location. To account for unequal sampling effort within each oceanic region, the relative abundance of sequences was standardized by dividing the total number of sequences of each order in each oceanic region by the number of samples within each oceanic region.

We performed Bray–Curtis similarity clustering and nMDS ordination to elucidate the differences in macroalgal assemblage among oceanic regions. One-way PERMANOVA and PERMDISP based on Bray–Curtis similarities were performed to test for differences in macroalgal assemblage composition across oceanic basins; data were log-transformed before these analyses. These analyses were performed in R using the Vegan[66] package. To evaluate how taxonomic richness and relative abundance of the sequences is distributed among oceanic regions, we calculated the indices of Pielou equitability (J), dominance (D) and Shannon (H); these indices assess evenness, dominance and diversity, respectively, at the order level. To compare observed order richness with estimated richness, we calculated the index CHAO2. Indices were calculated in PAST[67]. Order diversity was also evaluated through the water column from the surface to 4,000 m using only the Malaspina dataset; the Tara dataset was limited to a depth of 1,000 m.

The global distribution of macroalgal DNA sequences was analysed by assessing export and relative abundance with depth from the surface to the deep ocean (vertically), and with distance from the sampling point to the closest shoreline (horizontally). Vertical export was analysed by comparing Malaspina macroalgae sequences through the water column zones (that is, the epipelagic (0–200 m), mesopelagic (200–1,000 m) and bathypelagic zones (1,000–4,000 m)).

The attenuation of macroalgae sequences with depth was modelled by fitting the relative abundance of each zone to a normalized power function, following the coefficient[33] for particulate organic carbon flux:

$$y = ax^{-b}$$

where y is macroalgae relative abundance, a is the intercept, x is the depth and b is the macroalgae attenuation coefficient (sequences in reads per million per km). The relative abundance of macroalgal DNA by depth was standardized by dividing the total number of sequences per depth category by the number of samples within each depth category (0–200, 500, 1,000, 2,000, 3,000 and 4,000 m; the maximum depth recorded was 4,018 m).

Horizontal export was analysed by comparing the relative abundance of macroalgae sequences with the distance from the sampling point to the closest shoreline (continent or island) using one-way PERMANOVA. These data consisted of Tara and Malaspina metagenomes that belong to the epipelagic zone (0–200 m). The relative abundance of macroalgal DNA was standardized by dividing the total number of sequences per distance category by the number of samples within each distance category (0–200, 500, 1,000, 2,000, 3,000, 4,000 and 5,000 km; the maximum distance recorded was 4,860 km).

Data availability

The data that support the findings of this study can be found in ref. [20] (Tara Oceans metagenomes; https://doi.org/10.1038/sdata.2015.23), and ref. [57] (Tara Oceans 18S rDNA V9 metabarcodes; https://doi.org/10.1126/science.1261605) and Zenodo (Malaspina metagenomes; https://doi.org/10.5281/zenodo.2596829)[21].

References

50. Baldauf, S. L. The deep roots of eukaryotes. *Science* **300**, 1703–1706 (2003).
51. Zuccarello, G. C., Price, N., Verbruggen, H. & Leliaert, F. Analysis of a plastid multigene data set and the phylogenetic position of the marine macroalga *Caulerpa filiformis* (Chlorophyta). *J. Phycol.* **45**, 1206–1212 (2009).
52. Nakada, T., Misawa, K. & Nozaki, H. Molecular systematics of Volvocales (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic analyses. *Mol. Phylogenet. Evol.* **48**, 281–291 (2008).
53. Saunders, G. W. & Kucera, H. An evaluation of *rbcL*, *tufA*, UPA, LSU and ITS as DNA barcode markers for the marine green macroalgae. *Cryptogam. Algol.* **31**, 487–528 (2010).
54. Saunders, G. W. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Phil. Trans. R. Soc. Lond. B* **360**, 1879–1888 (2005).
55. Clerissi, C. et al. Unveiling of the diversity of *Prasinoviruses* (Phycodnaviridae) in marine samples by using high-throughput sequencing analyses of PCR-amplified DNA polymerase and major capsid protein genes. *Appl. Environ. Microbiol.* **80**, 3150–3160 (2014).
56. Pernice, M. C. et al. Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J.* **10**, 945–958 (2015).
57. De Vargas, C. et al. List of size fractionated eukaryotic plankton community samples and associated metadata (Database W1) *Pangaea* https://doi.org/10.1594/PANGAEA.843017 (2015).
58. Lanzén, A. et al. CREST—classification resources for environmental sequence tags. *PLoS ONE* **7**, e49334 (2012).
59. Cole, J. R., Konstantinidis, K., Farris, R. J. & Tiedje, J. M. in *Environmental Molecular Biology* (eds Liu, W.-T. & Jansson, J. K.) 1–20 (Horizon Scientific Press, 2010).
60. Giongo, A., Davis-Richardson, A. G., Crabb, D. B. & Triplett, E. W. TaxCollector: modifying current 16S rRNA databases for the rapid classification at six taxonomic levels. *Diversity* **2**, 1015–1025 (2010).
61. Hong, S.-H., Bunge, J., Jeon, S.-O. & Epstein, S. S. Predicting microbial species richness. *Proc. Natl Acad. Sci. USA* **103**, 117–122 (2006).
62. Schloss, P. D. & Handelsman, J. Status of the microbial census. *Microbiol. Mol. Biol. Rev.* **68**, 686–691 (2004).
63. Giner, C. R. et al. Marked changes in diversity and relative activity of picoeukaryotes with depth in the global ocean. Preprint at https://www.biorxiv.org/content/10.1101/552604v1 (2019).
64. Krause, L. et al. Taxonomic composition and gene content of a methane-producing microbial community isolated from a biogas reactor. *J. Biotechnol.* **136**, 91–101 (2008).
65. Huerta-Cepas, J. et al. EggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2015).
66. Oksanen, J. et al. vegan: Community ecology package. R package version 2.5-5 https://cran.r-project.org/web/packages/vegan/index.html (2007).
67. Hammer, Ř., Harper, D. A. T. & Ryan, P. D. PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **4**, 4 (2001).