Journal of
Marine Science and Engineering

MDPI

Review

Concepts towards Functional Eukaryotic Microbial Biogeography in the Ocean

Cora Hoerstmann [1,2,*], Sylke Wohlrab [2,3] and Uwe John [2,3,*]

[1] MIO—Mediterranean Institute of Oceanography, 13288 Marseille, France
[2] Alfred Wegner Institute Helmholtz Center for Polar and Marine Research, 27570 Bremerhaven, Germany
[3] Helmholtz Institute for Functional Marine Biodiversityat the University of Oldenburg, 23129 Oldenburg, Germany
* Correspondence: cora.hoerstmann@awi.de (C.H.); uwe.john@awi.de (U.J.)

Abstract: High-throughput sequencing technologies have revolutionized microbial diversity studies, shedding light on the oceans' plankton evolution, distribution, and biological activity. Whereas marine prokaryotes have been more extensively studied and specific methods developed, the research on microbial eukaryotes (protists) is falling behind, with major groups still largely unknown regarding their ecology and function. Because of numerous anthropogenic pressures, it is increasingly important to highlight the functional roles of protists in marine ecosystems. This review outlines the practices, challenges, and opportunities of high-throughput sequencing approaches (i.e., metabarcoding, metagenomics, and metatranscriptomics) to disentangle evolutionary, ecological, and functional aspects of protists in the ocean. These multidimensional approaches allow us to move from the classic picture of microbial biogeography towards functional microbial biogeography, explicitly highlighting the role of protists therein. We provide resources for functional classification and reflect on the current and future potential. We outline aspects of detecting and describing ecosystem changes at the species, population, and community levels, advancing methodological approaches for studying taxonomic diversity towards functional and evolutionary biodiversity concepts, seeking a more complete understanding and monitoring of ocean ecosystems.

Keywords: sequencing technologies; metabarcoding; metagenomics; metatranscriptomics

Citation: Hoerstmann, C.; Wohlrab, S.; John, U. Concepts towards Functional Eukaryotic Microbial Biogeography in the Ocean. *J. Mar. Sci. Eng.* **2022**, *10*, 1730. https://doi.org/10.3390/jmse10111730

Academic Editor: Linda Medlin

Received: 31 August 2022
Accepted: 7 November 2022
Published: 11 November 2022

1. Introduction

Microbial single-cell eukaryotic (protist) diversity in the ocean is rich, with an estimated total of 200,000–250,000 eukaryotic species [1]. Protists make up a significant fraction of the total microbial community filling fundamental roles in marine ecosystems and biogeochemical cycles. In marine environments, these small, planktonic organisms are continuously distributed through ocean currents and hydrodynamic processes and thus are exposed to changing environmental conditions. Therefore, protist diversity can be compared to a mosaic with different patterns at different temporal and spatial scales, breaking down into 'tiles' at the microscale [2,3].

The concept of microbial biogeography deals with the distribution of these protists and prokaryotes across multiple spatial scales and in regards to their adaptation over time [4]. Thus, importantly, it deals not only with the present distribution of microbial communities but also with the evolutionary processes driving these distributions. Further, there has been increasing interest in moving from taxonomy-focused studies, explaining communities and ecosystems by species lists, towards a more functional, trait-based understanding of ecosystems [5]. This is grounded in the fact that functional microbial patterns can differ from phylogenetic patterns through functional redundancy [6] and environmental factors driving both taxonomic diversity and functional diversity in different ways [7].

Changes in the functional diversity of protist communities therefore precede changes in biogeochemical cycles. Indeed, protist community composition can undergo irreversible

shifts in response to environmental conditions, such as temperature gradients [8] and, in particular, anthropogenic perturbations, such as climate change [9] and ocean acidification [10]. However, little is known about the major determinants of functional diversity, such as morphology, and the ecological roles of many taxonomic groups, despite their worldwide distribution [11]. These uncultured and nearly uncharacterised groups are referred to, for example, as marine alveolates (MALVs) [12], marine stramenopiles (MAST) [13], or marine haptophytes (MHAPTOs) [14]. These groups can contribute substantially to the overall molecular diversity in the nano- and picoplankton size fractions [15–17]; for example, heterotrophic organisms can account for a large proportion (up to more than 50%) of protists in coastal areas during the Arctic winter.

Extended global sampling efforts, such as those of Tara Oceans [18,19], Malaspina [20], and Ocean Sampling Day (OSD) [21], as well as a number of extended regional studies across spatial and temporal scales, have helped to identify endemic and regionally constrained species. For example, Supraha and colleagues [22] found distinct genotypes in the Arctic Ocean associated with different water masses and temperature regimes as well as cosmopolitan diatoms. The underlying genetic information of microorganisms reveals their individual roles within the ecosystem both in regards to functional overlaps between species (i.e., functional redundancy in the ecosystem) and individual species adaptations that allow the coexistence of multiple species at one site [23–25]. At each scale, protists play a crucial role in the ecosystem by providing resources, mediating biogeochemical cycles, and forming the base of the marine food web [26].

Specific tools are therefore needed to capture and assess functional protist diversity changes easily across spatial and temporal scales. High-throughput sequencing revolutionized the field of microbial diversity, because it allows the rapid (and more complete than microscopic) assessment of the general microbial diversity and biological processes [27,28]. Particularly, metabarcoding, metagenomics, and metatranscriptomics not only provide phylogenetic information about all microorganisms within one sample but are also increasingly recognized by their associated functional information (Figure 1). However, notably, these methods are often better adapted to the less complex prokaryotic cells, whereas protist diversity provides further challenges through the large genome sizes and a lack of reference material. The selection of an appropriate method targeting marker genes (metabarcoding), metagenomics, metatranscriptomics, or single-cell genomics depends on the objective of the research question to target the taxonomic and functional information of environmental microbial samples. Notably, other methods, such as microscopy and flow cytometry, can provide significant insights into the microbial trait space, as recently shown, for example, regarding the interactive effect of temperature and nutrients on phytoplankton communities [29]. Metabarcoding approaches in individual groups can also be used as specific molecular probes for other molecular techniques, such as fluorescence in situ hybridization (FISH). The connection between phylogenetic information from metabarcoding and the associated functional traits gained from trait tables or whole-genome sequencing (metagenomics, metatranscriptomics, and single-cell genomics) offers an effective way to add a functional perspective to ecosystem characterization, and therefore it may help elucidate the evolutionary and ecological processes shaping populations and communities and characterize their resilience capabilities in a changing world.

This review is intended to discuss the strength of the application of high-throughput sequencing approaches to marine protists to disentangle evolutionary, ecological, and functional aspects, focusing on protist diversity and function and moving from the classic picture of microbial biogeography towards functional microbial biogeography at large. This review outlines resources and approaches complementary to the more detailed review by Lopes dos Santos et al. [30] on the history and processing of metabarcode sequences. We focus on microbial research only, meaning living cells, although the application of eDNA (free-floating DNA) has also greatly advanced macroecology [31]. We highlight aspects of how to detect and describe ecosystem changes at the species, population, and community levels, moving from the "classical" metabarcoding approaches studying taxonomic diver-

sity toward functional and evolutionary biodiversity concepts, aiding a more complete understanding and monitoring of ocean ecosystems.

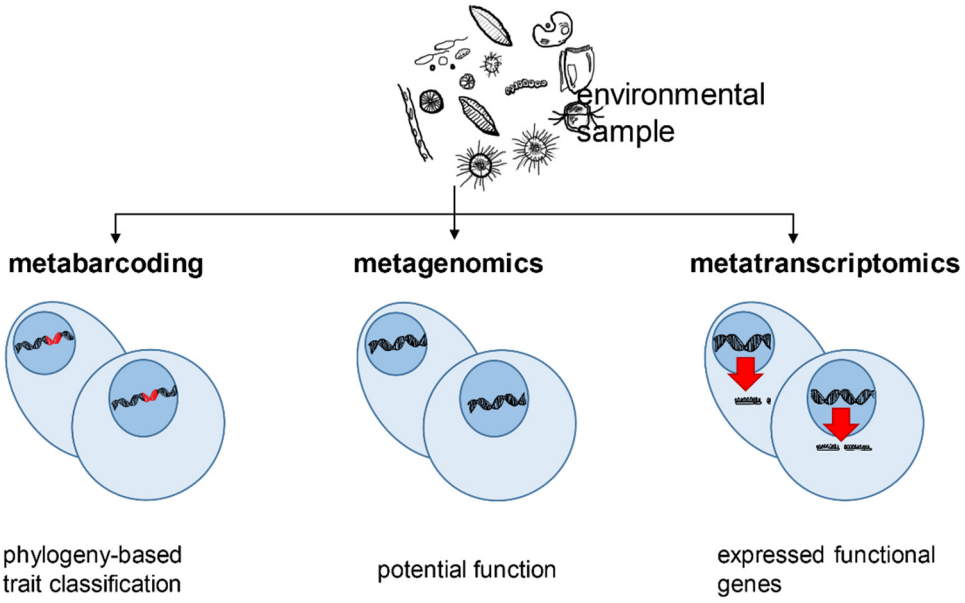

Figure 1. Schematic graph of DNA and RNA of an environmental sample accessed by metabarcoding, metagenomics, and metatranscriptomic approaches and their functional information.

2. High-Throughput Sequencing Approaches as an Evolutionary and Trait-Based Perspective on Phytoplankton Diversity

2.1. Metabarcoding

2.1.1. The Method

Metabarcoding in phyto- and mixoplankton [32] and for heterotroph protists is used to target specific taxonomic marker genes composed of hypervariable and evolutionary stable regions, such as the 16S, 18S, and 28S rRNA genes in the DNA. Metabarcoding is the principal method used to identify microbial communities in the ocean and has great global coverage [33]. The relatively short gene sequences (compared to whole genomes) offer relatively easy access to a deep understanding of the evolutionary profile of species; easy in the sense that we are able to access phylogenetic information at a lower cost (<USD 80/sample) than for whole-genome sequencing (>USD 200/sample) (Source: https://www.zymoresearch.com/blogs/blog/16s-sequencing-vs-shotgun-metagenomic-sequencing (accessed: 8 June 2022)) and less deep sequencing. As a result, one can extract operational taxonomic units (OTUs) or the nowadays more frequently used amplicon sequence variants (ASVs) that can be (if known) taxonomically annotated to the genera or species level, dependent on the reference database.

2.1.2. Challenges and Opportunities for Functional Diversity Analyses

The comparably straightforward approach of metabarcoding for identifying microbial diversity has led to its successful implementation in the field of marine microbiology at a higher spatial and temporal resolution than classical monitoring approaches. Indeed, we found 3–5 times more studies on marine metabarcoding than on marine metagenomics and 3–30 more than on metatranscriptomics, depending on the search engine (n = 3, Table S1). Therefore, metabarcoding data can be used for the monitoring and meta-analysis of multiple datasets. However, inherent shortcomings associated with metabarcoding data, methodological challenges, and limited data intercomparability complicate their application in the field.

In contrast to other high-throughput sequencing methods, metabarcoding includes a step wherein a specific gene of interest is amplified using dedicated primers that can identify the dataset's intercompatibility with the use of different primers. Efforts such as the

earth microbiome project have already helped to establish more standardized primers in the community, which have resulted in the widespread use of 18S rRNA primers proposed by Stoeck et al. [34] or modifications of these. These are the most widely used primers in marine samples. However, the limitation of this primer is that the 18S V4 (and even more the V9) regions are relatively short in sequence length. This provides only limited phylogenetic information and does not resolve all species' evolutionary histories [35]. Therefore, the use of other, more targeted primers is sometimes a better choice. For example, the 28S rRNA (the hypervariable D1 and D2 region) is useful as a target sequence for dinoflagellates [36,37], haptophytes [38], diatoms [39], and ciliates [40].

Metabarcoding is not a quantitative approach through the use of primers (Figure 2). On the one hand, the use of primers can be beneficial, as it can illuminate individual groups of interest. Even though read numbers are sometimes low, a high level of diversity can be revealed through the clustering into ASVs, such as the Rhizaria (Figure 3). On the other hand, general primers have biases against individual species, families, or even supergroups, such as the Excavata [41]. Therefore, it would be advisable, if possible, to cross-validate primers with the less biased technique of metagenomics, for example [35].

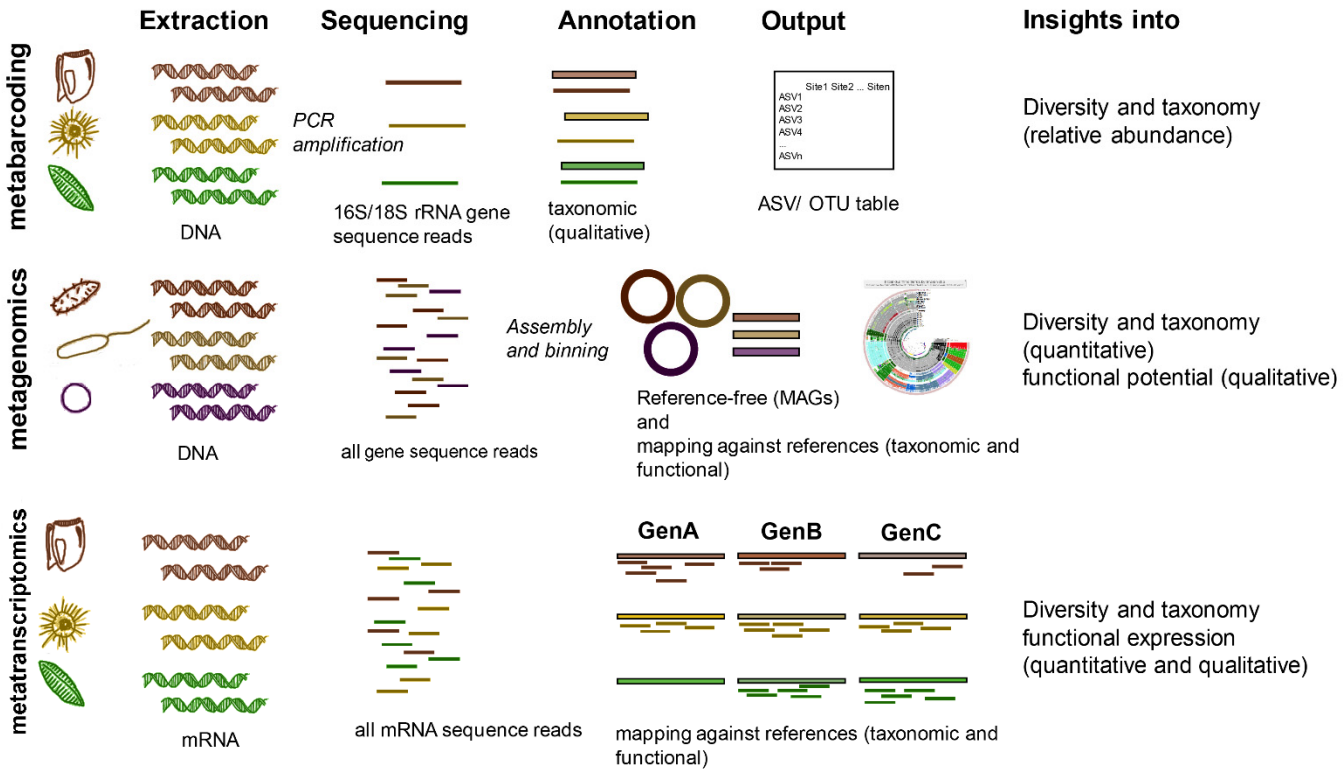

Figure 2. Schematic workflow and output of metabarcoding, metagenomics, and metatranscriptomics. Metagenomic output reflects potential visualization using the anvio platform.

The downstream processing of raw data in bioinformatic pipelines and the use of different reference databases bias the intercomparability of data and introduce uncertainties in meta-analyses. The FAIR data principles should counteract these limitations, but they are still not often applied for microbial metaomic data: among 635 studies of 16S rRNA metabarcoding, >65% contained data that were not available, not reusable, or contained faults in data formatting [42]. To increase the potential re-usability of metabarcoding data, multiple studies have been combined in the metaPR2 database, which allows a rapid cross-comparison of samples and the targeting of specific groups of interest [41]. This compilation of multiple datasets into one database is a significant step towards the global monitoring of microbial diversity.

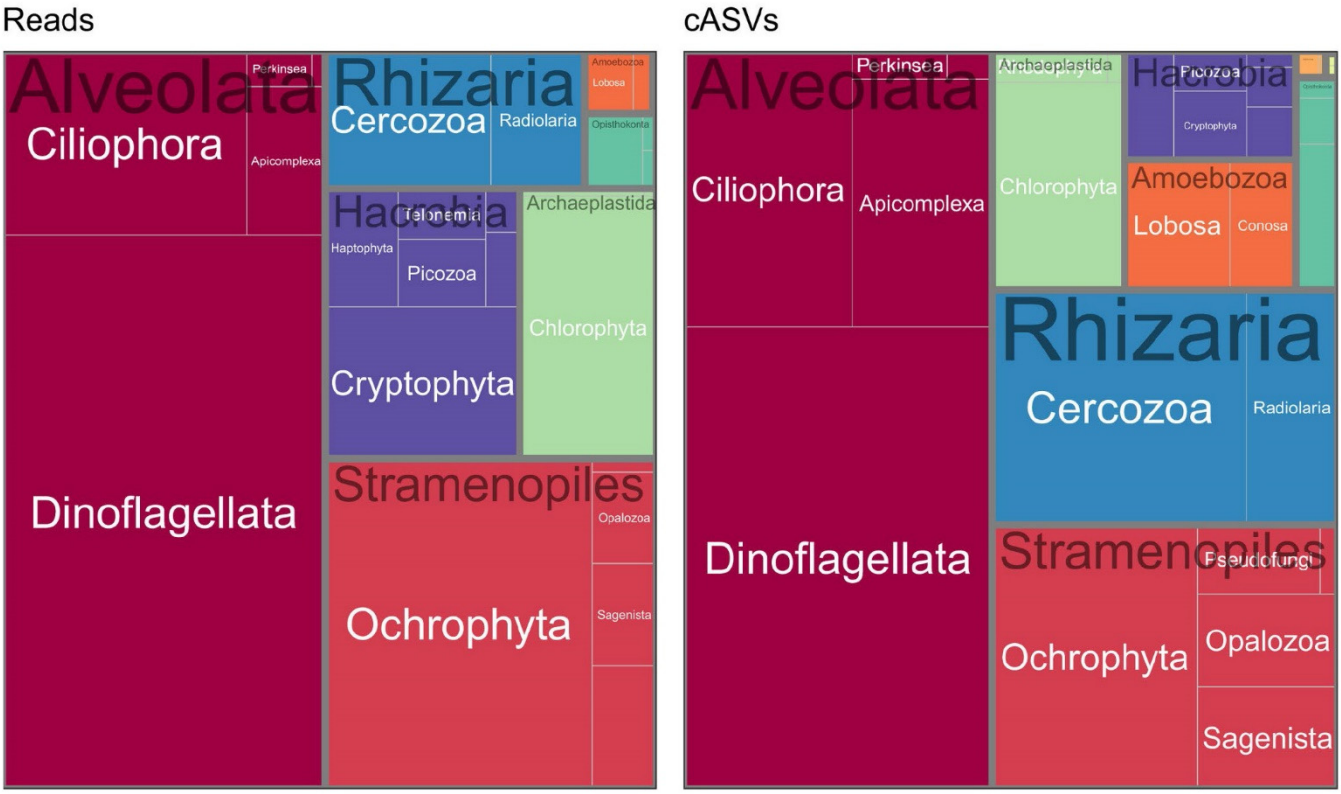

Figure 3. Treemaps of most abundant protist taxa (supergroups and divisions) for V4 datasets based on number of reads after normalization (**left**) or number of clustered ASVs (cASVs, (**right**)). Molecular Ecology Resources, volume: 22, issue: 8, pages: 3188–3201, first published: 28 June 2022, DOI: (10.1111/1755-0998.13674) [41].

The monitoring of microbial diversity can then be extended to the functional role of microbes in the ecosystem, accessed via knowledge about functional genes. Raes et al. [43] showed that in surface ocean samples across a large latitudinal transect in the Pacific Ocean, the bioinformatic tool 'PICRUSt2' works well to predict gene functions from 16S amplicon reads, which could potentially be similarly applied using the 18S rRNA [44]. Another possibility is the manual association of functional traits to known species through extensive literature research, as compiled in Table 1. Many trait-based analyses already exist [45], and their value has been repeatedly highlighted [46]. However, individual trait tables are rarely re-usable because of methodological challenges (e.g., data curation, repository maintenance) or the nature of the unique environment to which they are applied. For a more extensive overview, we summarized available trait tables (those we found and are aware of) that can be compiled and integrated into further analyses (Table 1). Such efforts would be most effective in the form of a compiled database for (marine) protist traits, as exists for freshwater phytoplankton (www.freshwaterecology.info, accessed on 6 November 2022), or in combination with the metaPR2 database. This approach would offer a useful simplification for improving models that use traits rather than taxonomic diversity (see, e.g., [47]).

Because metabarcoding is used to target specific genomic regions, it maximizes the genetic signal of specific (functional) groups, such as harmful algae [57,58], or DNA sequences that were previously difficult to access because of their low abundance in the environment or low sequence quality, including ancient DNA. For example, in their publication, Ibrahim et al. [59] describe how they used different metabarcoding markers targeting microbial eukaryotes, diatoms, and cyanobacteria to understand the processes driving protist plankton evolution in the light of paleoceanography. Thus, linking metabarcoding and paleoceanography with monitoring efforts will help us to understand not only the

changes towards the status quo of biodiversity but also future scenarios. This could also give rise to functional changes in past and future biogeochemical cycles, such as changes in carbon turnover in relation to primary production during ice-free events in the Arctic Ocean [60] or anthropogenic pollution [61]. Thus, the systematic monitoring of functional groups through metabarcoding can illuminate the importance of microbes' functional roles in the ecosystem, including their relevance to higher trophic levels [62].

Table 1. Existing phytoplankton trait tables [48–56].

Trait Table	Environment	Reference
Trophic mode	Marine (Arctic)	[48] https://doi.org/10.3897%2FBDJ.8.e56648
Trophic mode	Marine (Antarctic)	[49] https://doi.org/10.3389/fmicb.2022.844856
Trophic mode	Marine (Temperate)	[50] https://doi.org/10.1111/1462-2920.15832
Trophic mode	Freshwater	[51] https://doi.org/10.1111/fwb.12520
Large compilation of multiple traits and features (see machine-readable metadata)	Freshwater lakes (Temperate)	[52] https://doi.org/10.1038/s41597-021-00814-0
Freshwater ecology (www.freshwaterecology.info, accessed on 6 November 2022)	Freshwater (Temperate)	[53] https://doi.org/10.1016/j.ecolind.2015.02.007
Phytoplankton–zooplankton interactions	Freshwater	[54] https://doi.org/10.1007/s10750-015-2503-y
Trophic mode and morphology	Freshwater	[55] https://doi.org/10.1038/s41598-020-76645-7
Indicator and sensitivity values for saline lakes	Freshwater	[56] https://doi.org/10.1016/j.ecolind.2018.07.026

Despite the many advantages of metabarcoding, it remains a limited perspective on microbial communities and overlooks the functions and functional potential, which can only be more directly accessed through metagenomics and metatranscriptomics (Figure 2). Moreover, the inference of functional traits from taxonomy overlooks different ecotypes and pangenomes that are ultimately a core aspect of functional microbial biogeography. Metagenomics and metatranscriptomics have the great advantage of accessing the entire set of functional genes (metagenomics) and actually expressed functions (metatranscriptomics) within microbial communities.

2.2. Metagenomics

2.2.1. The Method

Metagenomic anlayses include the study of all genomic information within a sample. Whereas metabarcoding is an excellent tool for routine monitoring, observational studies, and even long-term programs, it only provides indirect information about the ecological and evolutionary processes of a community in response to environmental changes. In contrast, Dlugosh et al. [7] showed in an ocean metagenomics analysis that the functional profile of a microbial community responds more to biochemical variables, whereas the phylogenetic profile is more shaped by temperature and oceanographic provinces. This observation was also previously described in the South Indian Ocean [63], but the presence of functional redundancy could only be indirectly inferred based on primary productivity measurements and amplicon sequences.

2.2.2. Challenges and Opportunities for Functional Diversity Analyses

Metagenomics has the advantage that through the de novo assembly and binning of sequence reads, metagenome-assembled genomes (MAGs) can be detected and used to identify individual species without existing references (Figure 2). However, whereas metagenomics is well-established in other marine microbes, including prokaryotes and viruses, metagenomic studies of protists are still in their infancy because of the large genome size of many protists, the associated challenges for genome assembly, and the lack of references from cultures and single-cell genomics [64].

If appropriate gene references are available, metagenomics can shed light on species functions, which can be used to associate traits to species; to identify novel biocatalytic

activities, e.g., the degradation of toxic compounds [65]; and to explore different metabolic strategies for the same biological process (e.g., chitin degradation [66]). Therefore, using metagenomics, we not only gain insights into who is there and what they are doing, but also how they are doing it, allowing a further level of complexity that is relevant for microbial adaptations and resilience. Metagenomics can also be a tool to disprove links between species and traits that were previously thought to be conditional. For example, metagenomics analyses could show that *Trichodesmium* spp., considered one of the major contributors to marine N_2 fixation, appears to be non-diazotrophic under unfavorable low-oxygen conditions [67].

The direct association of genomic functions and environmental processes reveals relationships, such as the association between the presence of functional genes related to photosynthesis and growth in relation to an increase in carbon export, but it can also highlight the importance of marine viruses in mediating community dynamics in regards to carbon export efficiency [68]. Such network analyses and deep learning approaches pave the way towards a deeper understanding of these complex associations and more confident predictions.

2.3. Metatranscriptomics

2.3.1. The Method

Metatranscriptomes resolve the active part of the genomes of microeukaryotes in the current context of abiotic and biotic information processed by each cell through sequencing the mRNA of the organisms within a sample (Figure 2). By breaking down the relative abundance of genes by taxonomic groups, direct information can be obtained about why certain taxa are locally successful (reviewed in [69]). The method of metatranscriptomics therefore takes metagenomics a step further, because it can incorporate the dynamics of the biological response. Ecologically, the importance of gene expression can vary according to the environment, with changes in gene expression reflecting acclimation processes and/or species shifts, as shown by prokaryotic transcriptomes between temperate and polar regions [19]. Therefore, the study of metatranscriptomics adds both an internal (evolutionary adaptation) and external (functional within the community) temporal layer to the study of microbial functional biogeography. However, notably, metatranscriptomics does not reflect protein function. Therefore, only the combination of metagenomics, metatranscriptomics, and metaproteomics can reflect a microbe's functional potential and actions.

2.3.2. Challenges and Opportunities for Functional Diversity Analyses

Metatranscriptomics can reveal the biophysical and biochemical coupling between organisms and their environment, because it represents the most direct link between environmental conditions and biological responses of the three discussed methods. However, similarly to metagenomic analyses, metatranscriptomics is challenged by the number of individual transcriptomes within one sample and the lack of references. In addition, results obtained from metatranscriptomes are very sensitive towards modifications in sampling protocols as well as sampling time points.

Resolving the spatial and temporal variations in microbial activity provides a more comprehensive understanding of the influences on the qualitative and quantitative turnover of biogeochemical elements, such as carbon and nitrogen in the marine systems. For example, solar radiation could be associated with energy acquisition and metabolism that fuels dark respiration and the microbial loop, whereas membrane, amino acid, and vitamin biosynthesis occur at greater rates during the night [70,71]. Moreover, examining metatranscriptomics can be particularly useful in understanding the genes involved in responses to environmental stressors that help, for example, increase oil degradation rates [72]. On a spatial scale, Pearson et al. [73] suggest that >40% of diatoms' annotated expressed genes were community-specific, resulting in a functional landscape that could be associated with temperature, light, and nitrogen sources around the Antarctic Peninsula. In addition, metatranscriptomics revealed how dinoflagellate species adopt different trophic modes in

response to nutrient availability, from phototrophy to mixotrophy and heterotrophy [74]. Combining metatranscriptome data with machine learning extends further and can be applied to elucidate the trophic modes of whole communities, linking the flux of biogeochemical elements to ecology and environmental conditions [75]. This allows the refined analysis of adaptations to environments and helps researchers to understand and predict changes in ecosystem function.

Metatranscriptomics can also access information about evolutionary adaptations, because it is directly linked to selective environmental forces. Depending on gene expression, conserved and plastic community traits can be identified, which, on the one hand, contribute to the evolutionary understanding of the ecological success of a certain lineage and, on the other hand, allow researchers to identify traits that impact local changes in biogeochemical processes [72]. Moreover, co-occurrences and co-dependencies can be explored through gene expression patterns. For example, cell size and taxonomic affiliation can both play an important role in resource use and transfer within communities, and potentially in how adaptive processes operate in a changing environment [76,77]. In a temporal study tracing a regime shift from a diatom bloom to a dinoflagellate-dominated community, metatranscriptomics enabled the indentification of the metabolic and cellular evolutionarily manifested drivers of the respective taxa that provided them with their competitive advantage and allowed for bloom development [78]. Specifically, metatranscriptomics can provide insights into biochemical traits that taxonomic groups may inherit, and thus describe the dependencies of taxonomy and ecosystem processes on these functions as a testable hypothesis (Figure 4). In a comparative study [79], diatoms were shown to be an apparently abundant source of secondary metabolites from the mevalonate pathway (MVA) as compared to dinoflagellates. These include, for example, carotenoids that serve as a primary nutritional compound for multiple functions at higher trophic levels [80], which in turn may lead to adaptations to the expected availability of these compounds during the spring bloom. Such insights allow us to understand individual functional roles and species interactions, guiding us towards a more complete understanding of the drivers and ecological interactions within an entire ecosystem.

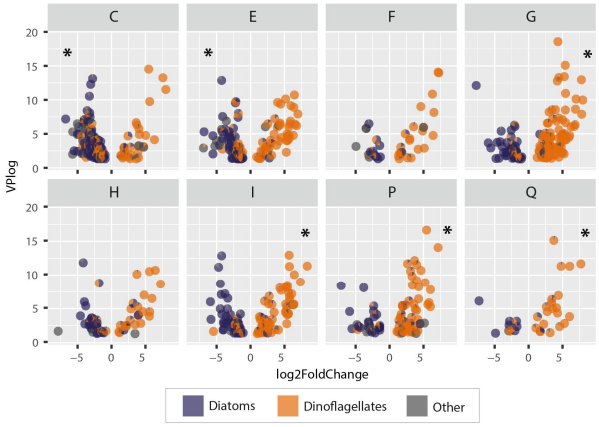

Figure 4. Differential gene expression in Greenland and Iceland coastal water samples. Negative values represent higher gene expression in Greenland samples; positive values represent higher gene expression in Iceland samples. Each point is illustrated by a pie chart representing a gene with its annotated relative taxonomic representation (diatoms, dinoflagellates, others). Each letter indicates a functional metabolic category; the names of the functional categories follow KOG: (C) energy production and conversion, (E) amino acid transport and metabolism, (F) nucleotide transport and metabolism, (G) carbohydrate transport and metabolism, (H) coenzyme transport and metabolism, (I) lipid transport and metabolism, (P) inorganic ion transport and metabolism, (Q) secondary metabolite biosynthesis, transport, and catabolism. Asterisks indicate a significant enrichment of genes in the respective category. Asterisks indicate a significant enrichment of expressed genes in the respective category. First published: 26 June 2020, DOI: 10.3389/fmars.2020.00439 [81].

3. Big Data in Metaomics

In parallel to the massive increase in marine microbial data products, advances in computational tools offer new avenues to analyze the big data in marine microbiology. The awareness regarding and application of these tools can be traced with a specifically developed u-index [82]. Whereas a review of these tools would go beyond the scope of this review, we would like to highlight the multiomics platform anvi'o (https://anvio.org/, accessed on 6 November 2022), which is built with multiple modular building blocks to establish workflows tailored for various research questions around metaomic data. Specifically, the visualization and interactive reconstruction of microbial genomes has enabled researchers to attain a high number of qualitative and refined metagenome-assembled genomes (MAGs), particularly for prokaryotes [83]. Generally speaking, as the possibilities of bioinformatic analysis advance, it is important that researchers are not overwhelmed by the amount and complexity of the available data, and that they have access to tools that allow them to explore more intuitively and creatively the complexity of microbial life in the ocean.

Whereas technologies have enabled us to access and explore the vastness of marine microbial diversity, we are still, to a large degree, 'in the dark' in regards to species diversity and genetic (and functional) potential, as well as in regards to exploring temporal and spatial coverage in the global oceans. Indeed, 40–70% of microbial genes lack reference and have unknown functions [84]. This is particularly true for the above discussed highly diverse and abundant groups, such as the dinoflagellates [85]. Additionally, there is generally more data available for prokaryotes than for protists, and protist ecology cannot simply be inferred from genomics, because of the more complex morphologies and behaviors of protists [64]. Moreover, sequencing-based approaches are more challenging for eukaryotic plankton because their genomes are large or gigantic (e.g., dinoflagellates and ciliates have genomes >100 × larger than the human genome) and exhibit expansive intergenic regions, many repeats, and only a very low percentage of coding regions [86,87].

4. What Do We Need for a Clear(er) Picture?

Multi-disciplinary, methodological, and analytical approaches will help close existing knowledge gaps in functional microbial eukaryotic microbiology. Whereas methodological advancements through, e.g., longer read sequencing, are exciting, it is as important to appreciate efforts to produce high-quality reference genomes as well as access and knowledge exchange within the research community. Indeed, the study of functional microbial eukaryotic biogeography is multidisciplinary itself, and only through well-structured and coordinated efforts will we be able to disentangle its complexity.

A significant challenge is that we still lack transcriptome and genome references to understand who is there and whose genes are being detected. A solution to these questions are compiled MAG databases, such as https://microbiomics.io/ocean/, accessed on 6 November 2022 [88] and the ocean DNA MAG catalog [89]. These can be used as a reference resource for metagenomics mapping. Moreover, the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) includes a database of 678 transcriptomes from cultures [90]. However, proportionally, much more data are available for prokaryotes than eukaryotes [64]. Thus, we must redress this bias and increase efforts to explore marine protists, particularly those whose functional roles are still unknown. Here, machine learning algorithms can help to infer trophic modes based on metatranscriptomic datasets [75]. This will provide new insights into protist ecology and function through the combination of advanced analytical methods and the increasing availability of reference data, which will also become easier in the future with new sequencing techniques.

An essential step in determining unknown functions and assembling genes into genomes and metagenome-assembled genomes (MAGs) is increasing the sequence read length. The third-generation sequencing systems, such as those of Pacbio or Nanopore (and soon Illumina), for instance, open a window onto larger molecular markers, allow us to sequence the entire ribosomal operon (18s-ITS1-5.8S-ITS2-28S rRNA), and will thus

help us in the future to identify species and different regional ecotypes more reliably when the reference database for the analyses have been adjusted. Furthermore, long-read sequence approaches, such as IsoSeq RNA, will generate mRNA sequences that can feature more functional information and thus provide the resources to facilitate the use of metatranscriptomic studies in many more ocean observational programs.

Standards and best practices are required to extend protist biogeography across groups, institutions, and programs towards global-scale observations that can disentangle the complex interplay of microbial organisms within one sample or region and across entire ocean basins. Specific hurdles include methodological biases, such as the use of different water volumes during sampling; different size fractionations (micrometers of filters); and the implementation of clean processing, avoiding the introduction of contaminations, which, at worst, lead to false conclusions that introduce errors in monitoring programs that persist over years or decades. Methodological developments, such as the use of mock communities in each sample [91], represent a practical solution to counter sampling, sequencing, and bioinformatics processing biases and to make samples more comparable, but post-processing and analytical methods that address the compositionality of the data are also important [92]. In a recent publication, Cohen et al. [93] compiled a number of studies and their metatranscriptomic workflows, allowing others to easily find suitable workflows for themselves. Moreover, the data need to be understandable to everyone. Here, the MIOP (minimum information of omic protocol) [94] helps creators of microbial sequence data understand what needs to be shared within the community. This can help remove redundancies (e.g., triplicates in PCR reactions [95]) and barriers to entering the field of microbial ecology.

5. Conclusions and a Way Forward

Multiple studies have shown that protist species patterns do not necessarily reflect a functional profile of individual geographic regions. Therefore, multiomics approaches are needed to disentangle these different patterns and to shed light on the link between protist diversity, their functional role within and across individual marine ecosystems, and their adaptations. This includes improving and expanding metabarcoding, metagenomic, and metatranscriptomic studies and interlinking the findings and insights from each of these approaches towards a complete understanding of protists' functional space. This also requires a sufficient level of spatial and temporal samples, which can only be achieved through concerted sampling efforts across working groups, initiatives, institutes, and countries. This would allow us to better understand and monitor ecosystem functions and states through a biological trait-based lens and could ultimately help us to better comprehend and react to the consequences of climate change and changes in species diversity, community assembly, and ecosystem function.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jmse10111730/s1.

Author Contributions: Conceptualization, C.H. and U.J., writing—original draft preparation, C.H., S.W. and U.J.; writing—review and editing, C.H., S.W. and U.J.; visualization, C.H.; funding acquisition, U.J. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received funding within the Helmholtz research program "Changing Earth, Sustaining our Future" (Sub-topic 6.2 Adaptation of marine life) of the Alfred-Wegener-Institut, Helmholtz Zentrum für Polar- und Meeresforschung, Germany.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Debroas, D.; Domaizon, I.; Humbert, J.F.; Jardillier, L.; Lepére, C.; Oudart, A.; Taib, N. Overview of freshwater microbial eukaryotes diversity: A first analysis of publicly available metabarcoding data. *FEMS Microbiol. Ecol.* **2017**, *93*, fix023. [CrossRef] [PubMed]
2. Seymour, J.R.; Amin, S.A.; Raina, J.B.; Stocker, R. Zooming in on the Phycosphere: The ecological interface for phytoplankton-bacteria relationships. *Nat. Microbiol.* **2017**, *2*, 17065. [CrossRef]
3. Mestre, M.; Höfer, J. The microbial conveyor belt: Connecting the globe through dispersion and dormancy. *Trends Microbiol.* **2021**, *29*, 482–492. [CrossRef] [PubMed]
4. Martiny, J.B.H.; Bohannan, B.J.M.; Brown, J.H.; Kane, M.; Krumins, J.A.; Kuske, C.R.; Morin, P.J.; Naeem, S.; Øvreås, L.; Reysenbach, A.; et al. Microbial biogeography: Putting microorganisms on the map. *Nature* **2006**, *4*, 102–112. [CrossRef] [PubMed]
5. Green, J.L.; Bohannan, B.J.M.; Whitaker, R.J. Microbial biogeography: From taxonomy to traits. *Science* **2008**, *320*, 1039–1043. [CrossRef]
6. Galand, P.E.; Pereira, O.; Hochart, C.; Auguet, J.C.; Debroas, D. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. *ISME J.* **2018**, *12*, 2470–2478. [CrossRef]
7. Dlugosch, L.; Poehlein, A.; Wemheuer, B.; Pfeiffer, B.; Badewien, T.H.; Daniel, R.; Simon, M. Significance of gene variants for the functional biogeography of the near-surface Atlantic Ocean microbiome. *Nat. Commun.* **2022**, *13*, 456. [CrossRef]
8. Martin, K.; Schmidt, K.; Toseland, A.; Boulton, C.A.; Barry, K.; Beszteri, B.; Brussaard, C.P.D.; Clum, A.; Daum, C.G.; Eloe-Fadrosh, E.; et al. The biogeographic differentiation of algal microbiomes in the upper ocean from pole to pole. *Nat. Commun.* **2021**, *12*, 5483. [CrossRef]
9. Williams, R.A.J.; Owens, H.L.; Clamp, J.; Peterson, A.T.; Warren, A.; Martín-Cereceda, M. Endemicity and climatic niche differentiation in three marine ciliated protists. *Limnol. Oceanogr.* **2018**, *63*, 2727–2736. [CrossRef]
10. Maas, E.W.; Law, C.S.; Hall, J.A.; Pickmere, S.; Currie, K.I.; Chang, F.H.; Voyles, K.M.; Caird, D. Effect of ocean acidification on bacterial abundance, activity and diversity in the Ross Sea, Antarctica. *Aquat. Microb. Ecol.* **2013**, *70*, ame01633. [CrossRef]
11. Forster, D.; Dunthorn, M.; Mahé, F.; Dolan, J.R.; Audic, S.; Bass, D.; Bittner, L.; Boutte, C.; Christen, R.; Claverie, J.M.; et al. Benthic protists: The under-charted majority. *FEMS Microbiol. Ecol.* **2016**, *92*, fiw120. [CrossRef] [PubMed]
12. Strassert, J.F.H.; Karnkowska, A.; Hehenberger, E.; Campo, J.; Kolisko, M.; Okamoto, N.; Burki, F.; Janou, J.; Poirier, C.; Leonard, G.; et al. Single cell genomics of uncultured marine Alveolates shows paraphyly of basal Dinoflagellates. *ISME J.* **2018**, *12*, 304–308. [CrossRef] [PubMed]
13. Massana, R.; Campo, J.; Sieracki, M.E. Exploring the uncultured microeukaryote majority in the oceans: Reevaluation of ribogroups within Stramenopiles. *ISME J.* **2014**, *8*, 854–866. [CrossRef]
14. Edvardsen, B.; Egge, E.S.; Vaulot, D. Diversity and distribution of Haptophytes revealed by environmental sequencing and metabarcoding—A Review. *Perspect. Phycol.* **2016**, *3*, 70176. [CrossRef]
15. Gran-Stadniczeñko, S.; Egge, E.; Hostyeva, V.; Logares, R.; Eikrem, W.; Edvardsen, B. Protist diversity and seasonal dynamics in Skagerrak plankton communities as revealed by metabarcoding and microscopy. *J. Eukaryot. Microbiol.* **2019**, *66*, 494–513. [CrossRef]
16. Massana, R.; Gobet, A.; Audic, S.; Bass, D.; Bittner, L.; Boutte, C.; Chambouvet, A.; Christen, R.; Claverie, J.M.; Decelle, J.; et al. Marine protist diversity in european coastal waters and sediments as revealed by high-throughput sequencing. *Environ. Microbiol.* **2015**, *17*, 4035–4049. [CrossRef]
17. Egge, E.; Elferink, S.; Vaulot, D.; John, U.; Bratbak, G.; Larsen, A.; Université, S.; De Roscoff, S.B. An 18S V4 rRNA metabarcoding dataset of protist diversity in the Atlantic inflow to the Arctic Ocean, through the year and down to 1000 m depth. *Earth Syst. Sci. Data* **2021**, *13*, 4913–4928. [CrossRef]
18. Pesant, S.; Not, F.; Picheral, M.; Kandels-Lewis, S.; Le Bescot, N.; Gorsky, G.; Iudicone, D.; Karsenti, E.; Speich, S.; Trouble, R.; et al. Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* **2015**, *2*, 150023. [CrossRef] [PubMed]
19. Salazar, G.; Paoli, L.; Alberti, A.; Huerta-Cepas, J.; Ruscheweyh, H.J.; Cuenca, M.; Field, C.M.; Coelho, L.P.; Cruaud, C.; Engelen, S.; et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* **2019**, *179*, 1068–1083.e21. [CrossRef]
20. Acinas, S.G.; Sánchez, P.; Salazar, G.; Cornejo-Castillo, F.M.; Sebastián, M.; Logares, R.; Royo-Llonch, M.; Paoli, L.; Sunagawa, S.; Hingamp, P.; et al. Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Commun. Biol.* **2021**, *4*, 604. [CrossRef]
21. Kopf, A.; Bicak, M.; Kottmann, R.; Schnetzer, J.; Kostadinov, I.; Lehmann, K.; Fernandez-guerra, A.; Jeanthon, C.; Rahav, E.; Ullrich, M.; et al. The Ocean Sampling Day Consortium. *GigaScience* **2015**, *4*, 27. [CrossRef]
22. Supraha, L.; Klemm, K.; Gran-stadniczenko, S.; Hoerstmann, C.; Vaulot, D.; Edvardsen, B.; John, U. Diversity and biogeography of planktonic Diatoms in Svalbard fjords: The role of dispersal and Arctic endemism in phytoplankton community structuring. *Elem. Sci. Anthr.* **2022**, *10*, 117. [CrossRef]
23. Roy, S.; Chattopadhyay, J. Towards a resolution of 'the Paradox of the Plankton': A brief overview of the proposed mechanisms. *Ecol. Complex.* **2007**, *4*, 26–33. [CrossRef]
24. Sauterey, B.; Ward, B.; Rault, J.; Bowler, C.; Claessen, D. The implications of eco-evolutionary processes for the emergence of marine plankton community biogeography. *Am. Nat.* **2017**, *190*, 116–130. [CrossRef] [PubMed]

25. Jahn, O.; Dutkiewicz, S.; Follows, M.J.; Ovidio, F.; Le, M. The dynamical landscape of marine phytoplankton diversity. *J. R. Soc. Interface* **2015**, 12. [CrossRef]
26. Delong, E.F.; Karl, D.M. Genomic perspectives in microbial oceanography. *Nature* **2005**, *437*, 336–342. [CrossRef] [PubMed]
27. Vaulot, D.; Eikrem, W.; Viprey, M.; Moreau, H. The diversity of small eukaryotic phytoplankton ($\leq$3 µm) in marine ecosystems. *FEMS Microbiol. Rev.* **2008**, *32*, 795–820. [CrossRef] [PubMed]
28. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Lozupone, C.A.; Turnbaugh, P.J.; Fierer, N.; Knight, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA.* **2011**, *108* (Suppl. 1), 4516–4522. [CrossRef]
29. Anderson, S.I.; Franzè, G.; Kling, J.D.; Wilburn, P.; Kremer, C.T.; Menden-Deuer, S.; Litchman, E.; Hutchins, D.A.; Rynearson, T.A. The interactive effects of temperature and nutrients on a spring phytoplankton community. *Limnol. Oceanogr.* **2022**, *67*, 634–645. [CrossRef]
30. Lopes dos Santos, A.; Gérikas Ribeiro, C.; Ong, D.; Garczarek, L.; Shi, X.L.; Nodder, S.D.; Vaulot, D.; Gutiérrez-Rodríguez, A. Phytoplankton diversity and ecology through the lens of high throughput sequencing technologies. In *Advances in Phytoplankton Ecology*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 353–413. [CrossRef]
31. Ruppert, K.M.; Kline, R.J.; Rahman, S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* **2019**, *17*, e00547. [CrossRef]
32. Flynn, K.J.; Maselli, M. Mixotrophic protists and a new paradigm for marine ecology: Where does plankton research go now ? *J. Plankton Res.* **2019**, *41*, 375–391. [CrossRef]
33. Burki, F.; Sandin, M.M.; Jamy, M. Diversity and ecology of protists revealed by metabarcoding. *Curr. Biol.* **2021**, *31*, R1267–R1280. [CrossRef]
34. Stoeck, T.; Bass, D.; Nebel, M.; Christen, R.; Meredith, D. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **2010**, *19*, 21–31. [CrossRef] [PubMed]
35. Latz, M.A.C.; Grujcic, V.; Brugel, S.; Lycken, J.; John, U.; Karlson, B.; Andersson, A.; Andersson, A.F. Short- and Long-Read Metabarcoding of the eukaryotic rRNA operon: Evaluation of primers and comparison to shotgun metagenomics sequencing. *Mol. Ecol. Resour.* **2022**, *22*, 2304–2318. [CrossRef] [PubMed]
36. John, U.; Cembella, A.; Hummert, C.; Elbrächter, M.; Groben, R.; Medlin, L. Discrimination of the toxigenic Dinoflagellates *Alexandrium Tamarense* and *A. Ostenfeldii* in cooccurring natural populations from Scottish coastal waters. *Eur. J. Phycol.* **2003**, *38*, 25–40. [CrossRef]
37. Medlin, L.; Elwood, H.J.; Stickel, S.; Sogin, M.L. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **1988**, *71*, 491–499. [CrossRef]
38. Bittner, L.; Egge, E.S. Diversity patterns of uncultured Haptophytes unravelled by pyrosequencing in Naples Bay. *Mol. Ecol.* **2013**, *33*, 87–101. [CrossRef]
39. Moniz, M.B.; Kaczmarska, I. Barcoding Diatoms: Is there a good marker ? *Mol. Ecol. Resour.* **2009**, *9*, 65–74. [CrossRef]
40. Stoeck, T.; Przybos, E.; Dunthorn, M. The D1-D2 region of the large subunit ribosomal DNA as barcode for Ciliates. *Mol. Ecol. Resour.* **2014**, *14*, 458–468. [CrossRef]
41. Vaulot, D.; Geisen, S.; Mahé, F.; Bass, D. Pr2-Primers: An 18S rRNA primer database for protists. *Mol. Ecol. Resour.* **2022**, *22*, 168–179. [CrossRef]
42. Jurburg, S.D.; Konzack, M.; Eisenhauer, N.; Heintz-buschart, A. The archives are half-empty: An assessment of the availability of microbial community sequencing data. *Commun. Biol.* **2020**, *3*, 474. [CrossRef] [PubMed]
43. Raes, E.J.; Karsh, K.; Sow, S.L.S.; Ostrowski, M.; Brown, M.V.; van de Kamp, J.; Franco-Santos, R.M.; Bodrossy, L.; Waite, A.M. Metabolic pathways inferred from a bacterial marker gene illuminate ecological changes across South Pacific frontal boundaries. *Nat. Commun.* **2021**, *12*, 2213. [CrossRef]
44. Douglas, G.M.; Maffei, V.J.; Zaneveld, J.R.; Yurgel, S.N.; Brown, J.R.; Taylor, C.M.; Huttenhower, C.; Langille, M.G.I. PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* **2020**, *38*, 669–673. [CrossRef]
45. Weithoff, G.; Beisner, B.E. Measures and approaches in trait-based phytoplankton community ecology—From freshwater to marine ecosystems. *Front. Mar. Sci.* **2019**, *6*, 40. [CrossRef]
46. Litchman, E.; de Tezanos Pinto, P.; Klausmeier, C.A.; Thomas, M.K.; Yoshiyama, K. Linking traits to species diversity and community structure in phytoplankton. *Hydrobiologia* **2010**, *653*, 15–28. [CrossRef]
47. Smith, S.L.; Vallina, S.M.; Merico, A. Phytoplankton size-diversity mediates an emergent trade-off in ecosystem functioning for rare versus frequent disturbances. *Sci. Rep.* **2016**, *6*, 34170. [CrossRef]
48. Schneider, L.K.; Anestis, K.; Mansour, J.; Anschütz, A.A.; Gypens, N.; Hansen, P.J.; John, U.; Klemm, K.; Martin, J.L.; Medic, N.; et al. A dataset on trophic modes of aquatic protists. *Biodivers. Data J.* **2020**, *8*, e56648. [CrossRef]
49. Grattepanche, J.-D.; Jeffrey, W.H.; Gast, R.J.; Sanders, R.W. Diversity of microbial eukaryotes along the west Antarctic Peninsula in austral spring. *Front. Microbiol.* **2022**, *13*, 844856. [CrossRef]
50. Hörstmann, C.; Buttigieg, P.L.; John, U.; Raes, E.J.; Wolf-Gladrow, D.; Bracher, A.; Waite, A.M. Microbial diversity through an oceanographic lens: Refining the concept of ocean provinces through trophic-level analysis and productivity-specific length scales. *Environ. Microbiol.* **2021**, *24*, 404–419. [CrossRef]
51. Salmaso, N.; Naselli-Flores, L.; Padisák, J. Functional classifications and their application in phytoplankton ecology. *Freshw. Biol.* **2015**, *60*, 603–619. [CrossRef]

52. Laplace-Treyture, C.; Derot, J.; Prévost, E.; Le Mat, A.; Jamoneau, A. Phytoplankton morpho-functional trait dataset from French water-bodies. *Sci. Data* **2021**, *8*, 40. [CrossRef] [PubMed]
53. Schmidt-Kloiber, A.; Hering, D. An Online Tool that Unifies, Standardises and Codifies more than 20,000 European Freshwater Organisms and Their Ecological Preferences. *Ecol. Indic.* **2015**, *53*, 271–282. [CrossRef]
54. Colina, M.; Calliari, D.; Carballo, C.; Kruk, C. A trait-based approach to summarize zooplankton–phytoplankton interactions in freshwaters. *Hydrobiologia* **2016**, *767*, 221–233. [CrossRef]
55. Borics, G.; B-Béres, V.; Bácsi, I.; Lukács, B.A.; T-Krasznai, E.; Botta-Dukát, Z.; Várbíró, G. Trait convergence and trait divergence in lake phytoplankton reflect community assembly rules. *Sci. Rep.* **2020**, *10*, 19559. [CrossRef] [PubMed]
56. Stenger-Kovács, C.; Körmendi, K.; Lengyel, E.; Abonyi, A.; Hajnal, É.; Szabó, B.; Buczkó, K.; Padisák, J. Expanding the trait-based concept of benthic diatoms: Development of trait- and species-based indices for conductivity as the master variable of ecological status in continental saline lakes. *Ecol. Indic.* **2018**, *95*, 63–74. [CrossRef]
57. Yarimizu, K.; Fujiyoshi, S.; Kawai, M.; Acuña, J.J.; Rilling, J.I.; Campos, M.; Vilugrón, J.; Cameron, H.; Vergara, K.; Gajardo, G.; et al. A standardized procedure for monitoring Harmful Algal Blooms in Chile by metabarcoding analysis. *J. Vis. Exp.* **2021**, *2021*, e62967. [CrossRef]
58. Jacobs-Palmer, E.; Gallego, R.; Cribari, K.; Keller, A.G.; Kelly, R.P. Environmental DNA metabarcoding for simultaneous monitoring and ecological assessment of many Harmful Algae. *Front. Ecol. Evol.* **2021**, *9*, 612107. [CrossRef]
59. Ibrahim, A.; Capo, E.; Wessels, M.; Martin, I.; Meyer, A.; Schleheck, D.; Epp, L.S. Anthropogenic impact on the historical phytoplankton community of Lake Constance reconstructed by multimarker analysis of sediment-core environmental DNA. *Mol. Ecol.* **2021**, *30*, 3040–3056. [CrossRef]
60. Pawłowska, J.; Wollenburg, J.E.; Zajączkowski, M.; Pawlowski, J. Planktonic Foraminifera genomic variations reflect paleoceanographic changes in the Arctic: Evidence from sedimentary ancient DNA. *Sci. Rep.* **2020**, *10*, 15102. [CrossRef]
61. Siano, R.; Lassudrie, M.; Cuzin, P.; Briant, N.; Loizeau, V.; Schmidt, S.; Ehrhold, A.; Mertens, K.N.; Lambert, C.; Quintric, L.; et al. Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. *Curr. Biol.* **2021**, *31*, 2682–2689.e7. [CrossRef]
62. Duffy, J.E.; Stachowicz, J.J. Why biodiversity is important to oceanography: Potential roles of genetic, species, and trophic diversity in pelagic ecosystem processes. *Mar. Ecol. Prog. Ser.* **2006**, *311*, 179–189. [CrossRef]
63. Hörstmann, C.; Raes, E.J.; Buttigieg, P.L.; Lo Monaco, C.; John, U.; Waite, A.M. Hydrographic fronts shape productivity, nitrogen fixation, and microbial community composition in the southern Indian Ocean and the Southern Ocean. *Biogeosciences* **2021**, *18*, 3733–3749. [CrossRef]
64. Keeling, P.J.; del Campo, J. Marine protists are not just big bacteria. *Curr. Biol.* **2017**, *27*, R541–R549. [CrossRef] [PubMed]
65. Kodzius, R.; Gojobori, T. Marine metagenomics as a source for bioprospecting. *Mar. Genom.* **2015**, *24*, 21–30. [CrossRef] [PubMed]
66. Raimundo, I.; Silva, R.; Meunier, L.; Valente, S.M.; Lago-Lestón, A.; Keller-Costa, T.; Costa, R. Functional metagenomics reveals differential chitin degradation and utilization features across free-living and host-associated marine microbiomes. *Microbiome* **2021**, *9*, 43. [CrossRef]
67. Delmont, T.O. Discovery of nondiazotrophic *Trichodesmium* species abundant and widespread in the open ocean. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2112355118. [CrossRef] [PubMed]
68. Guidi, L.; Chaffron, S.; Bittner, L.; Eveillard, D.; Larhlimi, A.; Roux, S.; Darzi, Y.; Audic, S.; Berline, L.; Brum, J.R.; et al. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **2016**, *532*, 465–470. [CrossRef] [PubMed]
69. Caron, D.A.; Alexander, H.; Allen, A.E.; Archibald, J.M.; Armbrust, E.V.; Bachy, C.; Bell, C.J.; Bharti, A.; Dyhrman, S.T.; Guida, S.M.; et al. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.* **2017**, *15*, 6–20. [CrossRef]
70. Poretsky, R.S.; Hewson, I.; Sun, S.; Allen, A.E.; Zehr, J.P.; Moran, M.A. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific Subtropical Gyre. *Environ. Microbiol.* **2009**, *11*, 1358–1375. [CrossRef]
71. Boysen, A.K.; Carlson, L.T.; Durham, B.P.; Groussman, R.D.; Aylward, F.O.; Ribalet, F.; Heal, K.R.; White, A.E.; DeLong, E.F.; Armbrust, E.V.; et al. Particulate metabolites and transcripts reflect diel oscillations of microbial activity in the surface ocean. *mSystems* **2021**, *6*, e00896-20. [CrossRef]
72. Song, B.; Li, Z.; Li, S.; Zhang, Z.; Fu, Q.; Wang, S.; Li, L.; Qi, S. Functional metagenomic and enrichment metatranscriptomic analysis of marine microbial activities within a marine oil spill area. *Environ. Pollut.* **2021**, *274*, 116555. [CrossRef] [PubMed]
73. Pearson, G.A.; Lago-leston, A.; Cánovas, F.; Cox, C.J.; Verret, F.; Lasternas, S.; Duarte, C.M.; Agusti, S.; Serrão, E.A. Metatranscriptomes reveal functional variation in Diatom communities from the Antarctic Peninsula. *ISME J.* **2015**, *9*, 2275–2289. [CrossRef] [PubMed]
74. Cohen, N.R.; McIlvin, M.R.; Moran, D.M.; Held, N.A.; Saunders, J.K.; Hawco, N.J.; Brosnahan, M.; DiTullio, G.R.; Lamborg, C.; McCrow, J.P.; et al. Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean. *Nat. Microbiol.* **2021**, *6*, 173–186. [CrossRef] [PubMed]
75. Lambert, B.S.; Groussman, R.D.; Schatz, M.J.; Coesel, S.N.; Durham, B.P.; Alverson, A.J.; White, A.E.; Armbrust, E.V. The Dynamic trophic architecture of open-ocean protist communities revealed through machine-guided metatranscriptomics. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2100916119. [CrossRef]

76. Wohlrab, S.; Falcke, J.M.; Lin, S.; Zhang, H.; Neuhaus, S.; Elferink, S.; Voss, D.; Zielinski, O.; John, U. Metatranscriptome profiling indicates size-dependent differentiation in plastic and conserved community traits and functional diversification in Dinoflagellate communities. *Front. Mar. Sci.* **2018**, *5*, 358. [CrossRef]
77. Alexander, H.; Jenkins, B.D.; Rynearson, T.A.; Dyhrman, S.T. Metatranscriptome analyses indicate resource partitioning between Diatoms in the field. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2182–E2190. [CrossRef]
78. Zhang, Y.; Lin, X.; Shi, X.; Lin, L.; Luo, H.; Li, L.; Lin, S. Metatranscriptomic signatures associated with phytoplankton regime shift from Diatom dominance to a Dinoflagellate bloom. *Front. Microbiol.* **2019**, *10*, 590. [CrossRef]
79. Elferink, S.; John, U.; Neuhaus, S.; Wohlrab, S. Functional genomics differentiate inherent and environmentally influenced traits in Dinoflagellate and Diatom communities. *Microorganisms* **2020**, *8*, 567. [CrossRef]
80. De Carvalho, C.C.C.R.; Caramujo, M.J. Carotenoids in aquatic ecosystems and aquaculture: A colorful business with implications for human health. *Front. Mar. Sci.* **2017**, *4*, 93. [CrossRef]
81. Elferink, S.; Wohlrab, S.; Neuhaus, S.; Cembella, A.; Harms, L.; John, U. Comparative metabarcoding and metatranscriptomic analysis of microeukaryotes within coastal surface waters of West Greenland and Northwest Iceland. *Front. Mar. Sci.* **2020**, *7*, 439. [CrossRef]
82. Callahan, A.; Winnenburg, R.; Shah, N.H. Analysis: U-index, a dataset and an impact metric for informatics tools and databases. *Sci. Data* **2018**, *5*, 180043. [CrossRef] [PubMed]
83. Eren, A.M.; Kiefl, E.; Shaiber, A.; Veseli, I.; Miller, S.E.; Schechter, M.S.; Fink, I.; Pan, J.N.; Yousef, M.; Fogarty, E.C.; et al. Community-led, integrated, reproducible multi-omics with Anvi'o. *Nat. Microbiol.* **2021**, *6*, 3–6. [CrossRef] [PubMed]
84. Vanni, C.; Schechter, M.S.; Acinas, S.G.; Barberán, A.; Buttigieg, P.L.; Casamayor, E.O.; Delmont, T.O.; Duarte, C.M.; Eren, A.M.; Finn, R.D.; et al. Unifying the known and unknown microbial coding sequence space. *eLife* **2022**, *11*, e67667. [CrossRef] [PubMed]
85. Stephens, T.G.; Ragan, M.A.; Bhattacharya, D.; Chan, C.X. Core genes in diverse Dinoflagellate lineages include a wealth of conserved dark genes with unknown functions. *Sci. Rep.* **2018**, *8*, 17175. [CrossRef]
86. John, U.; Lu, Y.; Wohlrab, S.; Groth, M.; Janouškovec, J.; Kohli, G.S.; Mark, F.C.; Bickmeyer, U.; Farhat, S.; Felder, M.; et al. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. *Sci. Adv.* **2019**, *5*, aav1110. [CrossRef]
87. Jaeckisch, N.; Yang, I.; Wohlrab, S.; Glöckner, G.; Kroymann, J.; Vogel, H.; Cembella, A.; John, U. Comparative genomic and transcriptomic characterization of the toxigenic marine Dinoflagellate *Alexandrium ostenfeldii*. *PLoS ONE* **2011**, *6*, e0028012. [CrossRef]
88. Paoli, L.; Ruscheweyh, H.-J.; Forneris, C.C.; Kautsar, S.; Clayssen, Q.; Salazar, G.; Milanese, A.; Gehrig, D.; Larralde, M.; Carroll, L.M.; et al. Uncharted biosynthetic potential of the ocean microbiome. *Nature* **2022**, *607*, 111–118. [CrossRef]
89. Nishimura, Y.; Yoshizawa, S. The OceanDNA MAG Catalog contains over 50,000 prokaryotic genomes originated from various marine environments. *Sci. Data* **2022**, *9*, 305. [CrossRef]
90. Keeling, P.J.; Burki, F.; Wilcox, H.M.; Allam, B.; Allen, E.E.; Amaral, L.A.; Armbrust, E.V.; Archibald, J.M.; Bharti, A.K.; Bell, C.J.; et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **2014**, *12*, e1001889. [CrossRef]
91. Yeh, Y.-C.; Needham, D.M.; Sieradzki, E.T.; Fuhrman, J.A. Taxon disappearance from microbiome analysis reinforces the value of mock communities as a standard in every sequncing run. *mSystems* **2018**, *3*, e00023-18. [CrossRef]
92. Gloor, G.B.; Macklaim, J.M.; Pawlowsky-Glahn, V.; Egozcue, J.J. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **2017**, *8*, 2224. [CrossRef] [PubMed]
93. Cohen, N.; Alexander, H.; Krinos, A.; Hu, S.K.; Lampe, R.H. Marine microeukaryote metatranscriptomics: Sample processing and bioinformatic workflow recommendations for ecological applications. *Front. Mar. Sci.* **2022**, *9*, 858. [CrossRef]
94. Samuel, R.M.; Meyer, R.; Buttigieg, P.L.; Davies, N.; Jeffery, N.W.; Meyer, C.; Pavloudi, C.; Pitz, K.J.; Sweetlove, M.; Theroux, S.; et al. Toward a global public repository of community protocols to encourage best practices in biomolecular ocean observing and research. *Front. Mar. Sci.* **2021**, *8*, 758694. [CrossRef]
95. Marotz, C.; Sharma, A.; Humphrey, G.; Gottel, N.; Daum, C.; Gilbert, J.A.; Eloe-Fadrosh, E.; Knight, R. Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary. *Biotechniques* **2019**, *67*, 29–32. [CrossRef]