frontiers

The osmolyte ties that bind: genomic insights into synthesis and breakdown of organic osmolytes in marine microbes

Erin L. McParland [1,*], **Harriet Alexander** [2,*] **and Winifred M. Johnson** [3,*]

[1] *Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA*
[2] *Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA*
[3] *Department of Chemistry and Biochemistry, University of North Carolina Wilmington, Wilmington, NC, USA*

Correspondence*:
Winifred Johnson
johnsonwm@uncw.edu

Harriet Alexander
halexander@whoi.edu
Erin L. McParland
emcparland@whoi.edu

ABSTRACT

~~Marine phytoplankton make up only .17ex1-2% of global plant biomass, yet they are responsible for approximately 40% of global carbon fixation. About 50% of this fixed carbon is consumed by heterotrophic bacteria each day, representing a large global flux of energy and matter~~The production and consumption of organic matter by marine organisms plays a central role in the marine carbon cycle. Labile organic compounds (metabolites) are the ~~primary supply of marine heterotrophs' energetic and nutritional demands~~ major currency of energetic demands and organismal interaction, but these compounds remain elusive because of their rapid turnover and concomitant minuscule concentrations in the dissolved organic matter pool. Organic osmolytes are a ~~class~~ group of small metabolites synthesized at high intracellular concentrations (mM) to regulate cellular osmolarity and have the potential to be released as abundant dissolved substrates. Osmolytes may represent an essential currency of exchange among heterotrophic prokaryotes and primary and secondary producers in marine food webs. For example, the well-known metabolite dimethylsulfoniopropionate (DMSP) is used as an osmolyte by some phytoplankton and can be subsequently metabolized by 60% of the marine bacterial community, supplying up to 13% of the bacterial carbon demand and 100% of the bacterial sulfur demand. While marine osmolytes have been studied for decades, our understanding of their cycling and significance within ~~the~~ microbial communities is still far from comprehensive.

Here, we surveyed the genes responsible for synthesis, breakdown, and transport of 14 key osmolytes. We systematically searched for these genes across marine bacterial genomes (n = 897) and ~~eukaryotic~~ protistan transcriptomes (n = 652) using homologous protein profiles to

investigate the potential for osmolyte metabolisms. Using the pattern of gene presence and absence, we infer the metabolic potential of surveyed microbes to interact with each osmolyte. Specifically, we identify: 1) ~~the potential to synthesize osmolytes~~ complete pathways for osmolyte synthesis in both prokaryotic and eukaryotic marine microbes, 2) ~~organisms that transport osmolytes but cannot synthesize or catabolize them~~microbes capable of transporting osmolytes but lacking complete synthesis and/or breakdown pathways, and 3) osmolytes whose synthesis and/~~breakdown~~ or breakdown appears to be specialized and is limited to a subset of organisms~~, i. e. niche-defining osmolytes.~~ . The analysis clearly demonstrates that the marine microbial loop has the genetic potential to actively recycle osmolytes and that this abundant ~~class~~ group of small metabolites may function as ~~micronutrients through requisite~~ a significant source of nutrients through exchange among diverse microbial groups that significantly contribute to the cycling of labile carbon.

Keywords: osmolytes, glycine betaine, mannitol, transporters, biosynthesis, catabolism, metatranscriptomics

1 INTRODUCTION

Marine microbes live in a high salinity environment and must maintain cellular mechanisms for rapid response to any small changes in external water potential in order to maintain cellular homeostasis (Bisson and Kirst, 1995). Ions are most readily transported by microbes, and therefore their concentrations typically respond first to external salinity changes. However, the charged nature of ions ~~mean that cellular physiology can only tolerate a specific concentration~~can only be tolerated by cellular physiology at specific concentrations, and ions are most often stored in solution within vacuoles in phytoplankton. Thus, organic osmolytes represent a critical addition to cellular osmotic balance. Throughout the rest of this paper when we refer to osmolytes, we will be exclusively discussing organic osmolytes. Osmolytes are often referred to as compatible solutes as they are small, organic molecules that are highly soluble and ~~are often referred to as compatible solutes. They~~ can accumulate at high concentrations within the cytoplasm without interfering with cellular function (Roeßler and Muller, 2001; Stefels, 2000). Although organic osmolytes ~~have a greater energy cost~~ are energetically more costly to use than inorganic ions, because they must be either synthesized or actively transported into the cell, many serve multiple functions in addition to their osmotic contributions. Osmolytes have been shown to contribute to protein stabilization and to have other cytoprotective properties under salinity, temperature, and pressure stress (Kirst, 1989; Ma et al., 2017; Yancey and Siebenaller, 2015; Burg and Ferraris, 2008). Some osmolytes ~~(dimethylsulfoniopropionate (DMSP), ectoine, proline, mannitol, and myo-inositol)~~ have been shown to be efficient scavengers of reactive oxygen species (Smirnoff and Cumbes, 1989; Sunda et al., 2002; Brands et al., 2019). ~~Finally, osmolytes~~ Osmolytes have also been predicted to function in a ballasting mechanism, where the synthesis of osmolytes with different densities would alter cellular buoyancy (Boyd and Gradmann, 2002; Lavoie et al., 2015). Osmolytes have been of interest in the marine context for decades (Challenger, 1951; Yancey et al., 1982; Kiene and Hoffmann Williams, 1998; Keller et al., 1999a). However, with the exception of a few extensively studied osmolytes (e.g. ~~DMSP~~dimethylsulfoniopropionate (DMSP)), we know relatively little about ~~their distribution~~ osmolyte distributions in the ocean, ~~how their intracellular abundance and resulting~~ the role of environmental conditions in controlling intracellular osmolyte abundances and subsequent release into the dissolved phase~~changes in response to environmental conditions, and their significance as a micronutrient~~ , or the significance of osmolytes as substrates in the microbial loop.

In marine systems, a range of amino acids and their derivatives, carbohydrates, methylsulfonium compounds, and methylammonium compounds have been identified as osmolytes used by marine organisms

ranging from bacteria to tube worms (Yancey, 2005) (Table 1). ~~Both~~ All macro- and microalgae exhibit ~~a general decrease in salinity tolerance as a function of isolation location where growth rates of open ocean isolates are more greatly impacted than that of coastal and or estuarine isolates (Brand, 1984; Kirst, 1989). Despite isolation location, all algae exhibited~~ the ability to accumulate both inorganic ions and organic osmolytes (Bisson and Kirst, 1995)~~. ~~, though their salinity tolerances have been shown to be a function of the isolation location's salinity conditions (open ocean vs. coastal ocean) (Brand, 1984; Kirst, 1989).~~ In general, phytoplankton contain one or two major organic osmolytes with concentrations $> 50mM$, and many minor organic osmolytes with concentrations $< 5mM$ (Gebser and Pohnert, 2013). However, monoculture studies have demonstrated that the composition of organic osmolytes is variable and strongly taxa dependent (Dickson and Kirst, 1987a,b). Marine phytoplankton use many well-known ~~osmolytes that have been studied extensively in plants~~ plant osmolytes (Yancey, 2005), but also osmolytes that are primarily found in the marine environment. For example, DMSP is found in a diverse array of marine phytoplankton, and was originally described as an osmolyte ~~, given~~ due to its similar structure to glycine betaine ~~, and has been widely studied relative to other osmolytes~~(Andreae, 1986). DMSP is ~~found in a diverse array of marine phytoplankton and is~~ hypothesized to be multi-functional, but likely functions predominantly as an osmolyte in dinoflagellate and haptophyte groups which typically contain $> 50mM$ intracellular DMSP (Keller et al., 1989; Stefels, 2000; McParland et al., 2020). 2,3-dihydroxypropane-1-sulfonate (DHPS) was very recently reported to be present at osmolyte-like concentrations in some diatoms ($15mM$~~max~~) and coccolithophores ($18mM$~~max~~) (Durham et al., 2019), and its intracellular concentrations in the Antarctic sea-ice diatom *Nitzschia lecointei* ($\sim 85mM$) correlate with salinity and temperature changes (Dawson et al., 2020a). Some marine diatoms have also been shown to be capable of transporting ~~DMSP or ectoine~~ ectoine or DMSP for osmotic functions (Spielmeyer et al., 2011; Lavoie et al., 2018; Fenizia et al., 2020). Across prokaryotes, halophilic archaea and bacteria have uniquely adapted cellular machinery that can withstand ~~the accumulation of~~ molar concentrations of potassium and chloride ions (Roeßler and Muller, 2001), but most marine prokaryotes also utilize organic osmolytes. Some prokaryotes maintain *de novo* synthesis pathways, while many others are capable of transporting osmolytes available in the dissolved pool (Poli et al., 2017; Poretsky et al., 2010). ~~The two~~ Two well-known marine nitrogen-fixing Cyanobacteria ~~contain two~~ use different organic osmolytes from each other: *Crocosphaera* synthesizes trehalose, while *Trichodesmium* synthesizes homoserine betaine (Pade et al., 2012, 2016). The chemoautotroph *Sulfurimonas* was recently shown to synthesize proline with increasing salinity (Götz et al., 2018), while *Vibrio parahaemolyticus* synthesizes ectoine as an osmolyte (Ongagna-Yhombi and Boyd, 2013). Rather than synthesizing osmolytes *de novo*, prokaryotes can also transport and retain ~~osmolytes~~ them under salinity stress. For example, uptake of radiolabeled DMSP and glycine betaine by natural communities of bacterioplankton has been shown to increase with salinity, and a significant proportion of the spiked osmolytes remained untransformed ~~within the particulate fraction~~ after uptake (Kiene and Hoffmann Williams, 1998; Motard-Côté and Kiene, 2015).

Osmolytes are some of the most concentrated (mM) small molecules in marine microbes, and therefore have the potential to be released as abundant carbon substrates to the dissolved organic matter (DOM) pool. The small size of osmolytes ~~likely make them labile sources of energy as they do not require additional extracellular processing in the way that larger biopolymers~~make them inherently labile substrates compared to larger biopolymers, such as proteins, polysaccharides, and lipids~~do~~, which require additional extracellular processing. Uptake studies of DMSP, glycine betaine, and taurine have estimated that these molecules ~~can~~ provide up to 13% (Kiene and Linn, 2000), 38% (Kiene and Hoffmann Williams, 1998), or 4% (Clifford et al., 2020) of bacterial carbon demand, respectively. When we consider the scale of the global carbon flux that is used by bacterioplankton (50% of net primary productivity (Ducklow, 2000)),

these substrates could contribute monumentally to this flux of energy and matter ~~. A rough estimate would suggest that given net primary productivity of $48.5 Pg$ $(10^{15} g)$ $C/year$ (Field et al., 1998), and assuming 50% of that carbon is required to fulfill bacterial carbon demand (Ducklow, 2000), individual osmolytes could have the potential to be cycled on the scale of $0.97 - 9.2 PgC/year$~~and dictate relationships amongst organisms. Catabolic pathways for other osmolytes, such as those identified for ectoine and 5-hydroxyectoine in the marine bacterium, *Ruegeria pomeroyi*, indicate that other osmolytes could also be valuable ~~and highly abundant~~ substrates (Schulz et al., 2017), but ~~we lack enough data about their concentrations in the environment to determine this~~lack relevant environmental measurements to quantify their abundance and role within the microbial loop.

Previous ~~observations~~ measurements of osmolytes within DOM range from $\sim pM - nM$ concentrations (Table 1), which contrast with the mM intracellular concentrations of osmolytes, suggesting that osmolytes are rapidly consumed upon release to the dissolved pool. As osmolyte concentrations can approach limits of detection with current analytical techniques, the down-regulation of transcripts associated with ~~the consumption of osmolytes~~ osmolyte consumption have been quantified as a proxy for utilization (Vorobev et al., 2018). When applying this technique to natural communities, the osmolytes glycine betaine, mannitol, taurine, proline, and sorbitol were all predicted to be rapidly recycled in less than 24 hours (Vorobev et al., 2018). Osmolyte ~~availability has~~ concentrations and availability have been hypothesized to be a function of both community composition and other environmental drivers. For example, DMSP concentrations were predicted to be determined by ~~community composition, where DMSP availability correlates with the distribution of phytoplankton species that contain $> 50 mM$ DMSP~~ phytoplankton community composition (McParland and Levine, 2019). Recent analytical advances have expanded our knowledge to a wide array of osmolytes and their environmental variability (including trehalose, sucrose, glucosylglycerol, DHPS, DMSP, glycine betaine, proline, homarine). Particulate abundances of some osmolytes are linked to diel cycles (Boysen et al., 2020), and different osmolytes dominate on sinking particles (glycine betaine, proline) compared to suspended particles (i.e. planktonic cells; DMSP) (Johnson et al., 2020). Although extensive measurements of DMSP concentrations and turnover rates exist (Kiene and Linn, 2000), the dissolved concentrations and turnover rates of most marine osmolytes are unknown (Table 1) and require further experimental work to fully understand the factors that ~~might be driving~~ drive their distributions.

In this study, we explored the role of osmolytes within the marine microbial loop by identifying microbes with the potential to synthesize, catabolize, assimilate, or transport osmolytes. We hypothesized that different patterns of osmolyte metabolisms would reveal the role of osmolytes ~~to serve as micronutrients~~ as valuable nutrients (i.e. potential carbon, nitrogen, and/or sulfur substrates) in the microbial loop. We first explored the taxonomic trends of potential producers and consumers by surveying osmolyte synthesis and utilization in genomes and transcriptomes from monocultures. We then compared these results with *in situ* transcription of mannitol and glycine betaine transport, synthesis, and breakdown in the Tara metatranscriptomes from the surface ocean.

2 MATERIALS AND METHODS

2.1 Manual curation of osmolyte-related KEGG orthologs

Osmolytes (n = 23) were initially selected for investigation based on their ubiquity within microorganisms or their previously identified importance in marine systems (Table 1). We chose to use the Kyoto Encyclopedia of Genes and Genomes (KEGG) as the database from which to identify metabolic pathways associated with the osmolytes and the corresponding gene orthologs, referred to as KEGG orthologs (KOs) throughout

this paper (Kanehisa and Goto, 2000; Kanehisa et al., 2012). Only 18 of these osmolytes of interest were ~~included in the~~ metabolites mapped in KEGG. ~~However~~In addition, glucosylglycerol had to be excluded because there were no genes associated with this compound in KEGG and trigonelline was also excluded due to missing KOs in the pathway available on KEGG. Dimethylsulfoniopropionate (DMSP) was also removed from the list of osmolytes considered because the pathways involved in its biosynthesis and breakdown are incomplete in KEGG. Finally, hydroxyectoine was not analyzed in detail as it is a closely linked derivative of ectoine. Thus, our analysis was performed on 14 osmolytes (Table 2).

The KOs associated with the synthesis, breakdown, and transport of these metabolites were manually compiled (Supplemental Data Sheet 1). We sought to identify synthesis pathways that began with a common metabolite, such as a sugar or amino acid, rather than an intermediate. Similarly, we mapped breakdown pathways to endpoints that could be incorporated back into metabolism for either catabolic or anabolic purposes. ~~While transporters~~ Transporters are not comprehensively represented in KEGG ~~, we selected those that were available . The annotation of osmolyte transporter systems in KEGG was limited to~~ and thus we chose to focus on the available ABC transporters. This limited us to transporters for glycerol, glycine betaine, mannitol, sorbitol, and taurine. By mapping the pathways in this way, we sought to identify organisms that could use ubiquitous building blocks to synthesize an osmolyte and those who could breakdown an osmolyte into biologically relevant molecules that could be used for another metabolic purpose.

To facilitate identification of synthesis and breakdown ~~capability~~ capabilities in organisms, we identified each of the genes required for a selected pathway by numerically labeling each gene as a step in a pathway as well as indicating alternative complete pathways (Supplemental Data Sheet 1). Finally, we created a comprehensive list of the KOs associated with synthesis, breakdown, and transport for each osmolyte to use in downstream homology searches. This resulted in 482 possible reactions or transporter components, but included replicate KOs because some pathways required the same KO as another pathway at certain steps. Ultimately, 486 unique KOs were identified (Supplemental Data Sheet 2). To confirm that our manual curation had comprehensively captured the genes associated with each osmolyte according to KEGG, we computationally searched a list of KOs directly associated with each osmolyte and checked this list against our manually curated pathways.

2.2 Homolog searching of KEGG orthologs against marine prokaryotic genomes and eukaryotic protistan transcriptomes

Genomes from representative marine bacteria and archaea were collected from MarRef (v5) (n = 897) (ten Hoopen et al., 2017; Klemetsen et al., 2017), and transcriptomes from representative marine protists were collected from the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) (n = 652) (Johnson et al., 2018; Keeling et al., 2014) (Supplemental Data Sheet 2). The predicted proteins from both the prokaryotic genomes and MMETSP transcriptomes were used in the subsequent analyses. All 486 KOs that were identified through manual curation as associated with the metabolic processing or transport of key osmolytes ~~2~~ (Table 2) were searched against the predicted proteins from the MMETSP and MarRef datasets. A snakemake pipeline (v. 5.24.2) (Mölder et al., 2021) based off of kofamscan (Aramaki et al., 2019; Mistry et al., 2013) was built to search for the likely presence of the osmolyte-related KOs in our protein sets of interest. KO hmm profiles were downloaded from KEGG (`https://www.genome.jp/tools/kofamkoala/`) on October 6, 2020. Both the MarRef and MMETSP predicted protein sets were searched for a given KO with hmmsearch (HMMER v. 3.3.1). True hits for KOs were parsed from the resulting hmm output with a custom script (parse_hmmsearch.py) run with python

v. 3.9.1 and biopython v. 1.78 (Cock et al., 2009), which parses hits based on the adaptive bitscore cutoff values used by `kofamscan` as published in the `ko_list` file (Aramaki et al., 2019). A final table with presence or absence of all osmolyte-related KOs for each organism was generated (Supplemental Data Sheet 2). This pipeline and all associated scripts and analysis notebooks is available as a GitHub repo (`https://github.com/AlexanderLabWHOI/2021-marine-osmolytes`).

2.3 Metabolic potential prediction

Using the manually curated pathway listings (Supplemental Data Sheet 1), a python based script was used to identify the presence or absence of a given metabolic pathway associated with the synthesis, breakdown, or transport for an organism (Supplemental Data Sheet 3). To be considered capable of osmolyte synthesis ~~and~~ or breakdown, an organism was required to have at least one complete pathway between the identified common metabolite (as described above). In ~~the case that~~ cases where a pathway had more than one step (Table 2), the organism had to have all steps present for it to be annotated as capable of synthesis or breakdown. In ~~the case that~~ cases where a step was identified on a pathway map, but was not annotated with a KO, the pathway was not considered. For the annotation of transporters, genomes were required to contain all KOs (subunits) of a given transporter to be annotated as capable of transport. In the case of glycine betaine, which contains two different transport pathways, only one transporter system was required to be present to confer the ability to transport.

2.4 Prokaryotic phylogeny comparison

Phylogenetic trees from the MarRef genomes were constructed using GToTree v1.4.7 (Lee, 2019; Price et al., 2010). Trees were constructed with the 'Universal' dataset of 15 genes as defined by Hug et al. (2016) and default parameters. Trees were visualized using iTOL (Letunic and Bork, 2016) and the presence or absence of synthesis, breakdown, and transport capabilities were visualized across all genomes and publication-quality figures were generated.

2.5 Identifying orthologs in Tara Oceans

To assess the potential importance of osmolyte metabolic processes across the global ocean, KOs from genes associated with glycine betaine and mannitol synthesis, breakdown, and transport were searched against the surface data from the Tara Oceans dataset using the Tara Oceans Gene Atlas (Villar et al., 2018). A selenium-based web bot was used to automate the search submission and result download process (`https://github.com/AlexanderLabWHOI/2021-marine-osmolytes/tree/master/genome-searching/tara-scraping`). KO file hmm profiles from kofamscan were uploaded to Tara Ocean gene atlas and searched against the metagenome and metatranscriptome datatsets for both the the 'eukaryotic' dataset, Marine Atlas of Tara Oceans Unigenes (MATOU v1) (Carradec et al., 2018), and Marine Atlas of Tara Oceans Unigenes and the 'prokaryotic' dataset, Ocean Microbial Reference Gene Catalog (OM-RGC v2, including the Arctic dataset) (Salazar et al., 2019). All searches were initially run with a generous e-value cutoff of 1e-10 and all count-based data were returned as "percent of mapped reads". We filtered the returned alignment files to retain only orthologs which passed the identified ~~bit-score~~ bitscore cutoff from kofamscan (Supplemental Data Sheet 4), such as we used in our genomic and transcriptomic analyses above. Notably, these cutoff values are quite stringent and significantly reduced the number of hits we reported (Supplemental Data Sheet 4). It is likely that assembled metagenomic or metatranscriptomic orthologs that are incomplete were not included in the analysis, but we

believe this conservative cutoff best reflects true hits in the Tara dataset. Taxonomic annotation of orthologs recovered was assessed using the reported taxonomy from Ocean Gene Atlas.

2.6 Plotting and statistical analysis

Data was analyzed and figures were generated with matplotlib v. 3.3.1 (Hunter, 2007), pandas v. 1.1.2 (pandas development team, 2020), and seaborn v. 0.11.0 (Waskom and the seaborn development team, 2020). Spearman correlations were calculated with the stats.spearmanr function within scipy v. 1.5.2 (Virtanen et al., 2020). Multi-testing p-value correction was done with the `multipletests` function within statsmodels 0.12.1 using the Bonferroni method. Global map distributions were plotted using cartopy 0.18.0 (Met Office, 2010 - 2015).

3 RESULTS AND DISCUSSION

3.1 Osmolyte synthesis, breakdown, and transport in marine prokaryote and eukaryote monocultures

In this study, we manually curated the genes responsible for the synthesis, breakdown, and transport of marine osmolytes (n = 14) (Table 1 & 2) that are classified as either amino acids and their derivatives (n = ~~9)or~~ 8), sugars (n = 5)~~(Table 2~~, or an amine oxide (n = 1). We then assessed the presence of these osmolyte-associated genes across marine prokaryotic genomes (MarRef) and marine eukaryotic transcriptomes from protists (MMETSP). The MarRef genomes (n = 897) contain representatives from 23 bacterial and 5 archaeal phyla. Notably, the genomes are biased towards copiotrophic lifestyles (and consequently more easily cultured and maintained in a laboratory setting) (Pachiadaki et al., 2019). The eukaryotic data targeted here constitute transcriptomes from protists, drawn from the MMETSP (n = 652). These transcriptome datasets contain representatives from across eight of the major supergroups within the eukaryotic tree of life.

Specifically, we looked at patterns of breakdown, synthesis, and transport across major prokaryotic and eukaryotic groups (Figure 1 & 2). The total number of pathways searched ranged from $1-75$ for synthesis and $1-50$ for breakdown (Table 2). The majority of pathways were comprised of a single step, though one breakdown (glutamate) and one synthesis pathway (ectoine) had a maximum 5 steps (Table 2). Relative to osmolyte synthesis and breakdown, ~~osmolyte transporters are under-characterized in KEGG~~ we looked at only a limited set of osmolyte transporters, focusing only on ABC transporters (n = 5 pathways total). Glycine betaine has two different transport systems annotated in KEGG, and the transporter for mannitol and sorbitol are identical (Supplemental Data Sheet 1). More than half of all prokaryotic genomes surveyed contained at least one of the transporter systems (n = 469), demonstrating the prevalence of transport in marine prokaryotes~~and the need for increased transporter annotation~~. The presence or absence of osmolyte synthesis, breakdown, and transport appeared to be phylogenetically linked in prokaryotes (Figure S1, S2 & S3).

We found that either: 1) breakdown and synthesis were tightly linked, suggesting core metabolism/internal recycling, or 2) breakdown and synthesis ability were unequal suggesting that the ~~osmolyte contributes to niche-forming metabolisms~~utilization of these osmolytes is a more specialized metabolism potentially only present in a smaller portion of the community. Additionally, most prokaryotic genomes that contained a transport system, also harbored the ability to synthesize and/or breakdown the respective osmolyte, except for glycine betaine for which many genomes were capable of transport without synthesis or breakdown.

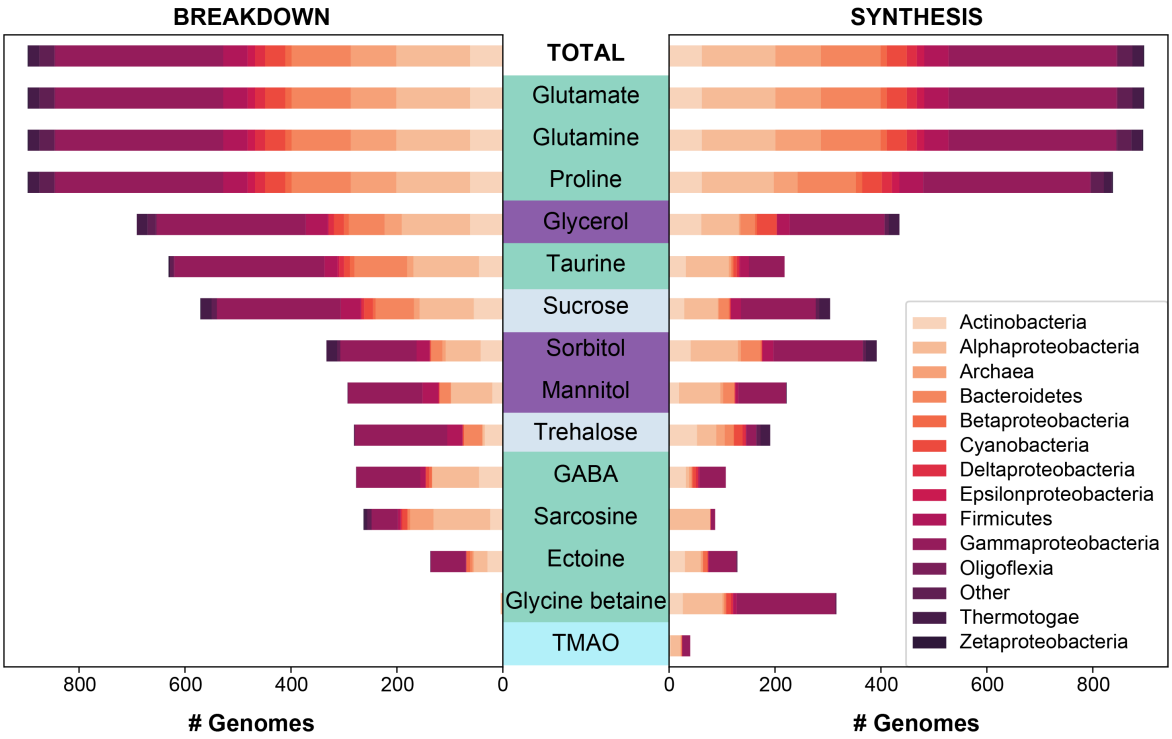

Figure 1. The predicted synthesis and breakdown of targeted osmolytes across all ('Total') MarRef bacterial and archaeal genomes (n = 897). The breakdown and synthesis of osmolytes is depicted as a stacked bar graph, and is colored by the designated taxonomic phylum (or class in the case of Proteobacteria). Osmolytes are colored by higher classification: amino ~~acid~~ acids and derivatives in green, sugar alcohols in purple, sugars in light periwinkle, and ~~amino~~ an amine oxide in cyan. Osmolytes are sorted along the y-axis based on the total number of genomes capable of breaking down a given osmolyte.

3.2 Amino acids and their derivatives

Glutamate and Glutamine

The ability to both synthesize and breakdown glutamate and glutamine was found in almost all prokaryotes (n = 897, n = 895, respectively) and eukaryotes (n = 647, n = 645, respectively) surveyed (Figure 3). This widespread function was expected as ~~not only do~~ these amino acids play both a fundamental role in the creation and structure of proteins, ~~but~~ and also in nitrogen ~~cycling~~ recycling in cells (Reitzer, 2003). Additionally, glutamate and glutamine abundance has been documented to rapidly change in response to altered osmotic conditions (Saum and Müller, 2008). Only two prokaryotic genomes lacked the ability to synthesize glutamine, which was likely due to either an incomplete genome assembly or an issue of homology in our searches. The synthesis and breakdown of glutamate and glutamine was also missing in just 3 and 4 eukaryotic transcriptomes, respectively. Again, this was likely due to lack of coverage in the transcriptome, and potentially related to the physiological condition of the organism at the time of sequencing.

Proline

Both breakdown and synthesis of proline was also common and widespread across prokaryotes (n = 838) and eukaryotes (n = 349), though slightly less so than glutamate and glutamine (Figure 3~~& ??~~). In particular, Archaea (52%) were less likely to have an identified proline synthesis pathway than bacterial groups (75-100%) (Figure 1). Proline synthesis was much more variable across eukaryotic supergroups,

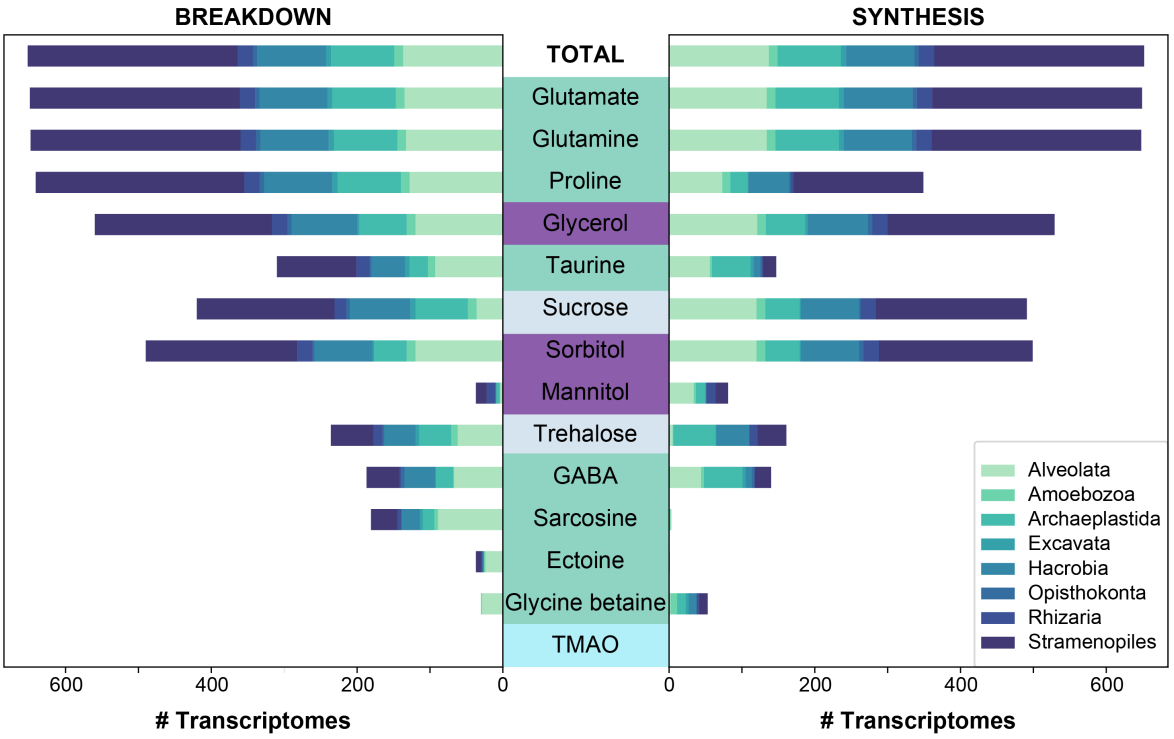

Figure 2. The predicted synthesis and breakdown of targeted osmolytes across all ('Total') MMETSP protist transcriptomes (n = 652). The breakdown and synthesis of osmolytes is depicted as a stacked bar graph, and is colored by the designated taxonomic supergroup. Osmolytes are colored by higher classification: amino ~~acid~~ acids and derivatives in green, sugar alcohols in purple, sugars in light periwinkle, and ~~amino~~ an amine oxide in cyan. Osmolytes are sorted along the y-axis based on the total number of genomes capable of breaking down a given osmolyte for bacteria and archaea to be consistent with Figure 1.

ranging from 14-92% (Figure 2). The observed trends for proline synthesis were similar to a previous study, which found that around 50% of bacterial genomes had a complete proline synthesis pathway, and only 40% and 30% of archaeal and eukaryotic genomes, respectively, were capable of synthesis (Mee and Wang, 2012). ~~In contrast, proline breakdown was detected in nearly every organism across prokaryotes and eukaryotes (Figure 3 & ??).~~

As with glutamate and glutamine, proline fulfills an important role as an amino acid in proteins, but is also an osmolyte for bacteria (Burg and Ferraris, 2008; Brill et al., 2011) and some marine diatoms (Dawson et al., 2020a,b), and can protect membranes from freezing (Yancey, 2005). More recently, proline was identified as the major organic osmolyte in the chemoautotroph *Sulfurimonas* found around hydrothermal vents (Götz et al., 2018). Interestingly, proline has been found to be abundant in deep sea particulate samples (Takasu and Nagata, 2015; Johnson et al., 2021) and in sinking marine particles (Johnson et al., 2020). One possible explanation for this could be increased use of proline as an osmolyte by organisms ~~adapting~~ adapted to the colder temperatures of the deep ocean.

Taurine

Broadly, the ability to synthesize taurine was less common than the ability to breakdown taurine across both prokaryotes and eukaryotes (Figure 3). A majority of prokaryotes were capable of taurine breakdown only (n = 442), and a smaller number were capable of both taurine breakdown and synthesis (n = 189) (~~??~~Figure 3). Taurine synthesis was most common within Alphaproteobacteria (58%) and Actinobacteria

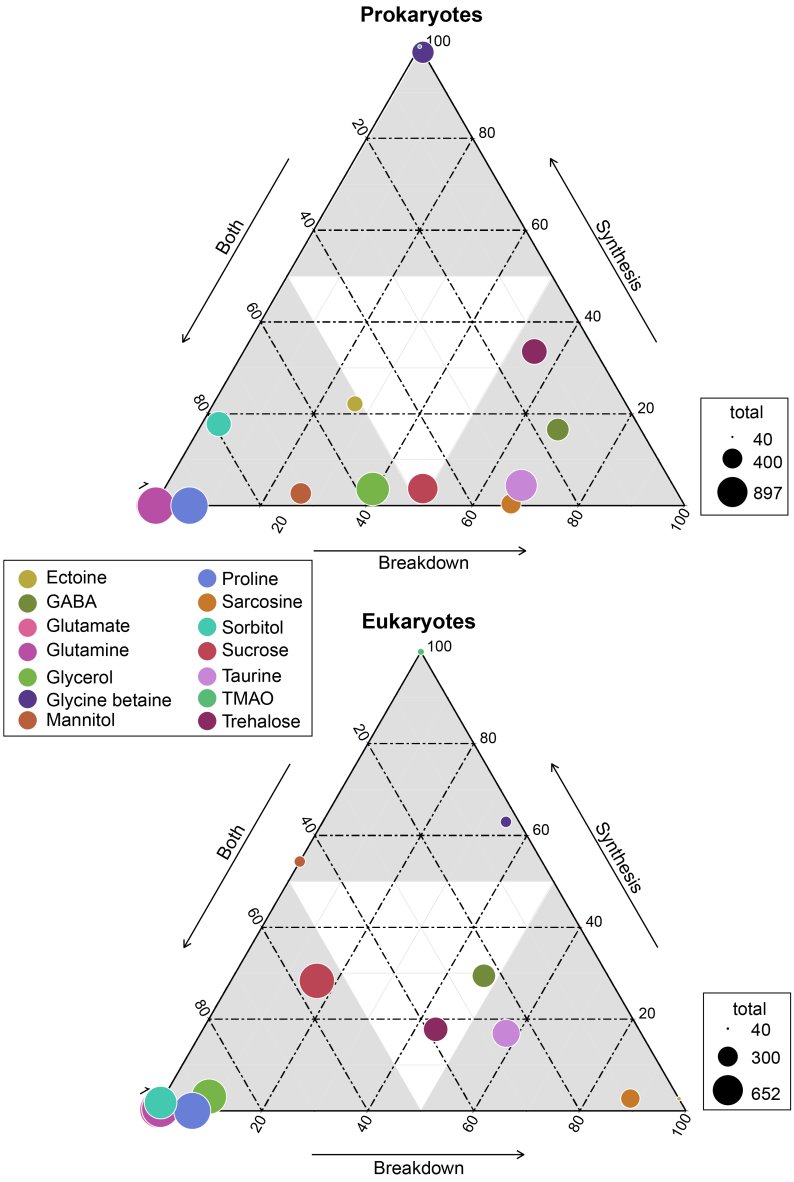

Figure 3. ~~The~~ Ternary plot of prokaryotes and eukaryotes osmolyte utilization based on genomic potential. Circle size represents total prokaryotes or eukaryotes capable of synthesis and/or breakdown of an osmolyte and circle placement represents the proportion of ~~predicted~~ the total capable of synthesis, breakdown, or both synthesis and breakdown. Circle sizes are scaled by the total number of ~~targeted osmolytes across all eukaryotic transcriptomes~~ MarRef genomes (n = ~~652~~897) and ~~all prokaryotic genomes~~ MMETSP transcriptomes (n = ~~897~~652) surveyed. Osmolytes are designated by color as depicted in the legend~~,~~. Glutamate is not clearly visible as it is found directly behind glutamine for both prokaryotes and ~~the groups (MMETSP and MarRef) are designated by shape (square and circle, respectively)~~eukaryotes.

(52%) (Figure 1). The breakdown of taurine in typically heterotrophic bacteria (i.e. Proteobacteria, Firmicutes, Bacteroidetes) ranged from 21% to 100% (Figure 1). Breakdown and synthesis of taurine was limited in Cyanobacteria (29% breakdown and 16% synthesis). These findings are consistent with previous work that identified bacterial uptake of taurine by three Proteobacteria groups: SAR11, Roseobacter, and *Alteromonas*, and also Thaumarchaeota and Euryarchaeota (Clifford et al., 2020, 2019). Metaproteomic data from the Ross Sea also indicated that SAR11 was a significant taurine sink via uptake and degradation

(Williams et al., 2012). Prokaryotes capable of taurine transport (n = 53) were almost exclusively limited to Proteobacteria and were all also capable of taurine synthesis and/or breakdown (Figure 4).

Taurine synthesis and breakdown was overall less common across eukaryotes (Figure 2 & 3). As observed for prokaryotes, the ability to only breakdown taurine was also most common in eukaryotes surveyed (n = 226) (??Figure 3). Taurine synthesis was most common in Archaeplastida (61%) (Figure 2). Interestingly, Archaeplastida were far less likely to be capable of taurine breakdown (29%). This contrasts with the other eukaryotic supergroups which had limited synthesis (e.g. ~~Stramenpiles~~ Stramenopiles (6%), Hacrobia, (10%), and Alveolata (41%)) and greater potential for breakdown (38%, 49%, and 68%, respectively) (Figure 2). Dissolved taurine concentrations in the Gulf of Alaska and the North Atlantic indicate that concentrations are typically in the low nanomolar range, and that release by amphipod-copepod assemblages in the Pacific occurs at rates of around 0.8 µmol g^{-1} C-biomass hr^{-1} and from Atlantic copepods at rates ranging from 1.3-9.5 μmol g^{-1} C-biomass hr^{-1} (Clifford et al., 2017). In the Adriatic Sea, even higher rates of release were found in mixed mesozooplankton communities, reaching the highest rates in the fall of an average of 59 μmol g^{-1} C-biomass hr^{-1} (Clifford et al., 2020). While our results indicate that protists are potential taurine sinks, zooplankton, specifically copepods, are clearly important taurine sources in the microbial loop. Multicellular eukaryotes (not represented in MMETSP) should be considered with respect to taurine recycling in the oceans, and possibly in addition to some other marine eukaryotes (mussels and tubeworms) that use taurine (Yin et al., 2000; Hosoi et al., 2005).

~~The co-occurrence of synthesis and breakdown within individual prokaryotic reference genomes from MarRef (top) and eukaryotic transcriptomes from the MMETSP (bottom). A bar graph depicts the predicted co-occurrence of synthesis and breakdown (black), only synthesis (turquoise), only breakdown (purple), or neither (gray) for each osmolyte surveyed across both the prokaryotic genomes (n = 897) and eukaryotic transcriptomes (n = 652).~~

Glycine Betaine

Glycine betaine synthesis was more common than breakdown in both prokaryotes and eukaryotes (Figure 3~~& ??~~). Most prokaryotes were only capable of glycine betaine synthesis (n = 316) (??Figure 3). In particular, a majority of Alphaproteobacteria (54%) and Gammaproteobacteria (59%) were capable of synthesis. The ability to breakdown glycine betaine was extremely rare across prokaryotes (Figure 1). The small number of prokaryotes capable of breakdown (n = 4) were not able to synthesize glycine betaine (??Figure 3) and were exclusively observed in Alphaproteobacteria strains: Candidatus *Pelagibacter ubique*, Alphaproteobacterium HIMB5, Alphaproteobacterium HIMB59, and *Rhodovulum sulfidophilum*. Glycine betaine synthesis and breakdown was also rare in Archaea, where only 4% were capable of synthesis and no Archaea were capable of breakdown. The ability to breakdown glycine betaine by bacteria is clearly a highly specialized metabolic ability. Interestingly, the majority of prokaryotes capable of glycine betaine transport cannot breakdown or synthesize (n = 190) or can only synthesize glycine betaine (n = 234) (Figure 4). This stands in stark contrast to our observations of taurine transporter co-occurrence with synthesis and/or breakdown, where no genomes contained only a taurine transporter (Figure 4). In addition, transport of glycine betaine was found more frequently than for other osmolytes~~. The~~ , and the presence of the two different glycine betaine transporters was taxa dependent. Proteobacteria primarily have the osmoprotectant ABC transporter, whereas Actinobacteria primarily have the glycine betaine/proline ABC transporter. A majority of Firmicutes contained both the osmoprotectant and the glycine betaine/proline ABC transporters, and notably did not contain any other osmolyte transporters (Figure 4). The widespread ability for glycine betaine transport without breakdown suggests that this osmolyte is likely targeted by bacteria for cellular retention to function as an osmolyte, which was observed previously by direct

measurements of radiolabeled glycine betaine uptake (Kiene and Hoffmann Williams, 1998). Glycine betaine was the only osmolyte in our survey that appeared to be targeted for osmotic function by bacteria, rather than as a carbon and/or nitrogen substrate to be metabolized.

Glycine betaine synthesis and breakdown was also limited in eukaryotes (Figure 2), and the majority were only capable of synthesis (n = 53) (??Figure 3). Notably, glycine betaine synthesis was most common in Amoebozoa (75%) and Excavata (57%), while in other groups, synthesis was $< 14\%$ prevalent (Figure 2). The breakdown of glycine betaine was absent from all groups except for Alveolata (21%) and Amoebozoa (8%) (Figure 2). Glycine betaine was previously measured in monocultures of Hacrobia (*Chrysochromulina sp.*, *Emiliania huxleyi*) and Alveolata (*Amphidinium carteraea*, *Prorocentrum minimum*), and some diatoms (*Thalassiosira pseudonana*) (Gebser and Pohnert, 2013; Keller et al., 1999a,b). These direct ~~osmolyte measurements~~ measurements of cellular osmolytes suggest we under-predicted glycine betaine synthesis in our analysis. It is possible however that glycine betaine synthesis was not upregulated in the conditions the reference transcriptomes were collected in.

Sarcosine

Sarcosine is structurally very similar to glycine betaine, but it lacks two methyl groups relative to glycine betaine. Sarcosine breakdown was much more common in both prokaryotes and eukaryotes (Figure 3~~& ??~~). Sarcosine synthesis was not present in prokaryotes, except Alphaproteobacteria of which 55% contained a synthesis pathway (Figure 1). Relative to synthesis, sarcosine breakdown was far more common across prokaryotic groups, particularly in Actinobacteria, Archaea, and Alphaproteobacteria (39%, 51%, 77%, respectively). Sarcosine synthesis was almost completely absent in all eukaryotes (Figure 2). As observed for prokaryotes, more eukaryotes were capable of sarcosine breakdown, including most Amoebozoa, Excavata, and Alveolata (42%, 57%, 65%, respectively). While sarcosine as an osmolyte in marine organisms does not seem to have been extensively studied, elasmobranchs have been shown to use sarcosine to counteract increased levels of urea (Treberg et al., 2006), and, generally, for decades, sarcosine has been considered to act as an osmolyte (Arakawa and Timasheff, 1985). Sarcosine can be synthesized from creatine, choline, or glycine (Supplemental Data Sheet 1). As the ability to synthesize sarcosine was mostly absent in prokaryotic and eukaryotic groups, except in Alphaproteobacteria, it suggests that the source of sarcosine in the ocean is relatively unknown. However, sarcosine is an intermediary product formed during the breakdown of glycine betaine (glycine betaine $\rightarrow$dimethylglycine $\rightarrow$sarcosine $\rightarrow$glycine) (Supplemental Data Sheet 1), and therefore the incomplete breakdown of glycine betaine could be a possible source of sarcosine in the ocean.

Ectoine

Ectoine metabolism was less widely distributed across prokaryotes and eukaryotes compared to the other osmolytes (Figure 1 & 2). Within the small number of prokaryotes with ectoine metabolism, most were capable of both synthesis and breakdown (n = 90), while a smaller number were only capable of breakdown (n = 47) or only capable of synthesis (n = 39) (??Figure 3). Ectoine synthesis and breakdown had a fairly similar taxonomic distribution within prokaryotes. For example, $\sim 48\%$ of Actinobacteria and $\sim 20\%$ of Alphaproteobacteria were capable of both synthesis and breakdown (Figure 1). However, the Betaproteobacteria (n = 12 genomes in MarRef) had the highest proportion of synthesis (75%) and breakdown (58%). Notably, ectoine synthesis and breakdown pathways were completely absent in Cyanobacteria, and were very rare in Archaea (5% synthesis, 2% breakdown). Ectoine is known to be an important osmolyte for *Halomonas* sp. (Ono et al., 1999) and has also been found to be used across species of *Vibrio*, both those that are associated with other organisms, like fish or shellfish, or planktonic

species (Pflughoeft et al., 2003; Ongagna-Yhombi and Boyd, 2013; Ma et al., 2017). Ectoine synthesis was completely absent in eukaryotes, except for one Hacrobia strain, which was likely a result of prokaryotic contamination of the transcriptome (Figure 2~~& ??~~). A very small number of eukaryotes were capable of ectoine breakdown ($< 18\%$ in all supergroups). Ectoine is typically considered to be a prokaryotic osmolyte, but was recently observed in Stramenopiles (n = 2 diatoms), Hacrobia (n = 2 haptophytes, n = 1 coccolithophore), and Alveolata (n = 1 dinoflagellate) (Fenizia et al., 2020). Specifically, the diatoms were shown to take-up ectoine in xenic cultures, and also synthesize ectoine in axenic monocultures, but only one of the three genes used by bacteria for ectoine synthesis had a putative homolog in the genome of the diatom *Phaeodactylum tricornutum* (Fenizia et al., 2020). This low homology with bacterial ectoine synthesis genes may explain why we also did not find significant ectoine synthesis in the eukaryotic transcriptomes surveyed here.

GABA

The synthesis and breakdown of Gamma-aminobutyric acid (GABA) was also found to be relatively rare compared to other osmolytes in both prokaryotes and eukaryotes (Figure 3). Most prokaryotes were only capable of GABA breakdown (n = 225), compared to a smaller number capable of synthesis only (n = 55), or both synthesis and breakdown (n = 52) (~~??~~Figure 3). A majority of Actinobacteria (73%) and Alphaproteobacteria (63%) were capable of GABA breakdown (Figure 1). GABA synthesis was also common in Actinobacteria (52%), but less so in other groups (e.g. Alphaproteobacteria 4%). The ability to only breakdown GABA was most common in eukaryotes (n = 125) (~~??~~Figure 3), but some proportion of every supergroup was capable of GABA synthesis or breakdown, ranging from 5% - 61% synthesis and 8% - 83% breakdown across supergroups (Figure 2). In particular, a large portion of Archaeplastida were capable of synthesis (57%), but breakdown was relatively constrained (27%). GABA has been identified in certain species of marine yeast at much higher concentrations than commercial yeast (Masuda et al., 2008), and is an indicator of ecosystem health in snapper (Goode et al., 2020). In marine bacteria, greater quantities of GABA are released from cells in response to decreased sodium chloride concentrations or increased pH (Mountfort and Pybus, 1992). GABA is also used as a settlement queue for marine invertebrates, however, around 1/3 of bacteria isolated from a potential settlement site were also found to metabolize GABA and have high- and low-affinity transporters for the substrate (Kaspar et al., 1991). This suggests a complex role in marine environments where GABA serves as a signaling compound, osmolyte, and carbon substrate.

3.3 Amine oxide

TMAO

Only one pathway for synthesis and breakdown of TMAO was identified within the KEGG framework (Table 2), which limited our analysis of TMAO synthesis, breakdown, and transport as more pathways likely exist. We identified synthesis pathways for trimethylamine N-oxide (TMAO) in a few prokaryotic groups: Actinobacteria (5%), Alphaproteobacteria (14%), Bacteroidetes (2%), Cyanobacteria (3%), and Gammaproteobacteria (5%). However, this is likely due to limitations of the pathways annotated in KEGG. TMAO is found in deep sea animals including teleosts, skates, and crustaceans where it is thought to play an additional role as a protectant from the high hydrostatic pressure of the deep ocean (Yancey et al., 2002). Nanomolar concentrations of TMAO have been measured in Antarctic surface waters as well, indicating a role throughout the water column (Gibb and Hatton, 2004), and copepods have been shown to produce TMAO from trimethylamine (Strom, 1979).

3.4 Sugars and sugar alcohols

Sorbitol

Both the breakdown and synthesis of sorbitol were equally distributed in prokaryotes and eukaryotes, though a larger percentage of eukaryotes had the ability to breakdown and synthesize sorbitol (Figure 3). More than 40% of all Actinobacteria, Firmicutes, Alphaproteobacteria, and Gammaproteobacteria were capable of sorbitol breakdown and synthesis (Figure 1). Cyanobacteria ~~were not capable of sorbitol breakdown or synthesis~~ appeared to be missing all of the single step pathways for sorbitol breakdown and sorbitol synthesis (Table 2 & Figure 1). Sorbitol synthesis by prokaryotes was more common than expected as prokaryotes typically do not use polyols (e.g. sorbitol and mannitol) as compatible solutes (Kinne, 1993; Empadinhas and Costa, 2008). Both sorbose and sorbitol were previously identified as metabolites that are rapidly consumed from the dissolved pool (Vorobev et al., 2018). Here, the breakdown of sorbose to sorbitol was annotated as a sorbitol synthesis pathway, and therefore it is possible that many of the prokaryotes identified to be capable of sorbitol synthesis actually utilize this pathway to breakdown sorbose. Additionally, the majority of prokaryotes that were capable of sorbitol transport (n = 100) also contained sorbitol breakdown and synthesis pathways (Figure 4). Only a small number of prokaryotes (n = 7) contained all subunits of the sorbitol/mannitol transporter without the ability to synthesize or breakdown sorbitol. Sorbitol synthesis and breakdown was broadly and commonly distributed across the eukaryotic supergroups surveyed (Figure 2). An average $\sim 70\%$ of each eukaryotic group was also capable of sorbitol breakdown and/or synthesis, except for the Excavata (Figure 2). The widespread ability to synthesize and breakdown sorbitol in eukaryotes was expected as polyols have previously been observed to function as compatible solutes in eukaryotes (Burg and Ferraris, 2008), and sorbitol has been observed to increase with salinity in microalgae (Brown and Hellebust, 1978).

Mannitol

Mannitol breakdown and synthesis was less common in both prokaryotes and eukaryotes (Figure 3). Mannitol synthesis and breakdown was broadly found across all Proteobacteria ($< 60\%$) and Actinobacteria ($\sim 30\%$), whereas more Firmicutes had the potential to breakdown mannitol (65%) compared to synthesis (11%) (Figure 1). Mannitol was also not expected to be widespread across prokaryotes as polyols are uncommon in prokaryotes (Empadinhas and Costa, 2008), and mannitol synthesis has been previously observed in prokaryotes but in a limited number of taxa. Specifically, a soil Gammaproteobacterium synthesized mannitol *de novo* when other osmolyte sources were not provided exogenously (Sand et al., 2013). As observed for sorbitol, few prokaryotes (n = 26) ~~contained the~~ were capable of sorbitol/mannitol transport ~~function without the ability to synthesize or breakdown mannitol (Figure 4). A~~ only, and the majority of prokaryotes capable of mannitol transport were also able to both synthesize and breakdown mannitol (n = 124) (Figure 4). The presence of the transporter alongside breakdown and synthesis pathways suggests that 1) sorbitol and mannitol are not transported for osmotic function only, and 2) the transporter may actually be important for efflux with respect to internal recycling of the osmolytes. Mannitol utilization was even less common in eukaryotes (Figure 2). Eukaryotic synthesis and breakdown of mannitol was most common in Rhizaria ($> 50\%$), but not common in their potential Hacrobia symbiotic partners (1%)~~(Figure 2). Mannitol is most commonly considered~~ . Mannitol was previously found to be present in Archaeplastida (Kirst, 1989), and was shown ~~specifically~~ to increase with salinity in three ~~Chlorophyte species~~ Archaeplastida strains of Chlorophytes (Dickson and Kirst, 1987b). Relative to Rhizaria though, mannitol synthesis was limited ~~to~~ in Archaeplastida (only 15%~~of Archaeplastida~~), though this supergroup encompasses a large diversity of eukaryotes.

Glycerol

Glycerol synthesis and breakdown was common across prokaryotes and eukaryotes, but the presence of breakdown and synthesis was not even in prokaryotes (Figure 3~~& ??~~). While Actinobacteria were almost equally capable of synthesis (98%) and breakdown (100%), Proteobacteria were 2-fold more capable of breakdown (93%) than synthesis (51%) (Figure 1). In contrast, Cyanobacteria were 2-fold more capable of synthesis (92%) than breakdown (50%). Globally, 31% of prokaryotes were only capable of glycerol breakdown, lacking the ability to synthesize (~~??~~Figure 3). Prokaryotes that can transport glycerol (n = 139) were almost exclusively limited to Proteobacteria and are all capable of synthesis and/or breakdown of glycerol (Figure 4). In eukaryotes, 78% were capable of both synthesis and breakdown of glycerol, with the ability broadly spread across groups (Figure 2 & ~~??~~3). Glycerol is one of the major byproducts of photosynthesis that is often released and shared within algal-host symbioses and has also been found to be induced under osmotic stress in model symbiotic algae such as *Symbiodinium* (Suescún-Bolívar et al., 2016; Mayfield and Gates, 2007). Moreover, glycerol has been found to increase the activity and abundance of mixotroph-associated bacteria (e.g. Alphaproteobacteria and Gammaproteobacteria), and may act as a key currency of exchange between protists and prokaryotes in the marine system (Poddar et al., 2018).

Sucrose

The breakdown of sucrose was at least 2-fold more common than the synthesis of sucrose in all prokaryotic groups, except for Thermotogae (Figure 3~~& ??). The bias towards~~). All of the references in the Thermotogae group surveyed in MarRef (n = 20) are capable of both synthesis and breakdown of sucrose. This group includes known hyperthermophilic chemoorganotrophic organisms

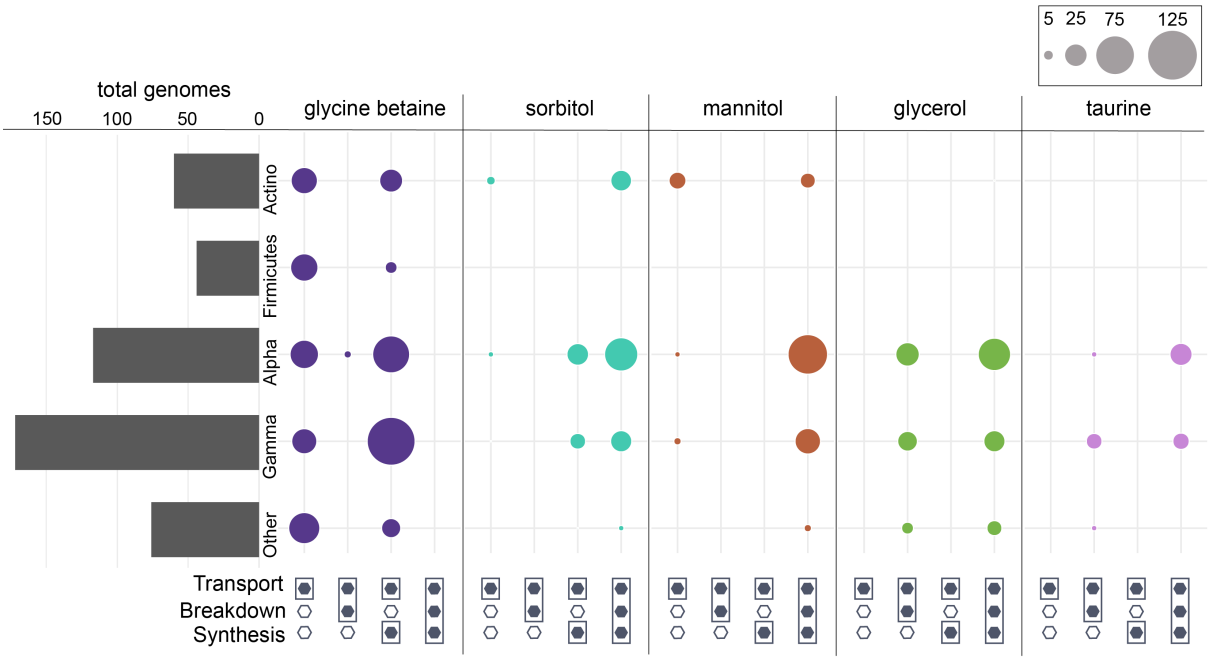

Figure 4. Prokaryotic genomes that contained one or more of the ~~annotated~~ osmolyte ABC transporters annotated in KEGG. Bars represent the total number of genomes that contained at least one transporter. 'Other~~group~~' represents taxa that contributed $< 10\%$ of the total. For each respective osmolyte, circles represent the total number of genomes in each group that were capable of transport only, transport and breakdown, transport and synthesis, or all (transport, breakdown, and synthesis). Missing circles indicate no genomes with the respective category.

such as *Thermotoga maritima*, which are known to metabolize many simple and complex carbohydrates (Nelson et al., 2001). The general bias towards sucrose breakdown may suggest ~~these groups~~ that prokaryotes primarily rely on other organisms for sucrose supply. ~~Notably,~~ Although sucrose has been shown to serve an osmotic function in Cyanobacteria and other photosynthetic organisms (Reed et al., 1986; Klähn and Hagemann, 2011), the sucrose synthesis pathway was unexpectedly incomplete based on our three-step definition (Supplemental Data Sheet 1) ~~, the expected sucrose synthesis pathway was incomplete~~ in Cyanobacteria genomes surveyed in MarRef (n = 38), which included *Synechococcus*, *Prochlorococcus*, known nitrogen fixers, and others. ~~This stands in contrast to evidence suggesting that sucrose specifically serves an osmotic function in Cyanobacteria and other photosynthetic organisms (Reed et al., 1986; Klähn and Hagemann, 2011).~~ All Cyanobacteria genomes were missing the pathway's first step (conversion of glucose-1-phosphate to UDP-glucose). The majority of *Synechococcus* and *Prochlorococcus* genomes contained the glucosyltransferase required for the second step (conversion of UDP-glucose to sucrose-6-phosphate), but ~~the phosphohydrolase~~ not the phosphatase required for the third step (sucrose-6-phosphate hydrolysis)~~was absent~~. The few other Cyanobacteria (*Calothrix* and *Nodularia*) (n = 9) did contain the phosphatase for the third step but were missing the enzymes for the first two steps. ~~All of the references in the Thermotogae group surveyed in MarRef~~ Sucrose utilization was common in eukaryotes (n = ~~20)are capable of synthesis and breakdown of sucrose. This group includes known hyperthermophilic chemoorganotrophic organisms such as *Thermotoga maritima*, which is known to metabolize many simple and complex carbohydrates (Nelson et al., 2001).~~

~~Compared to prokaryotes, sucrose synthesis (~ 80%) in eukaryotic references was slightly more common than breakdown (~ 65%), with 50% of surveyed protists~~ 586), and the majority were capable of both synthesis and breakdown (Figure 3~~& ??~~). Amoebozoa, Hacrobia, Rhizaria, and Stramenopiles were all most frequently capable of both synthesis and breakdown (Figure 2). In contrast, Alveolata had 3-fold more synthesis pathways than breakdown~~(Figure 2). This may indicate lack of a key gene in the breakdown pathway or potentially that sucrose is also~~ , suggesting that sucrose may potentially be a significant osmolyte for dinoflagellates.

Trehalose

The breakdown and synthesis of trehalose exhibited unique trends across the major prokaryotic groups surveyed, where either only synthesis or only breakdown were more common than the ability to both synthesize and breakdown (Figure 1 & ~~??~~3). This suggests that trehalose is less likely to be internally recycled, but rather it is either synthesized as an osmolyte, or consumed as a carbon source. Trehalose synthesis was more common in Actinobacteria (86%), Alphaproteobacteria (26%), Archaea (19%), and Cyanobacteria (42%), whereas trehalose breakdown was more common in Bacteriodetes (30%), Firmicutes (63%), and Gammaproteobacteria (55%) (Figure 1). Interestingly, 90% of Thermotogae surveyed were capable of trehalose synthesis, but no trehalose breakdown. Trehalose was found to be a primary osmolyte in *Crocosphaera watsonii*, instead of glucosylglycerol which is typically used by marine Cyanobacteria (Pade et al., 2012). Many non-marine bacteria use this osmolyte, suggesting that this capability in *Crocosphaera* may have been obtained through horizontal gene transfer (Pade et al., 2012). A new ~~sequence of the~~ trehalose synthase gene that converts maltose to trehalose has also been identified in a marine species of *Pseudomonas* (Gao et al., 2013). Intracellular concentrations of trehalose have been found to fluctuate on a diel cycle in the north Pacific subtropical gyre (Boysen et al., 2020), suggesting that it might be important in the physiology of oligotrophic bacteria, such as nitrogen fixers. Trehalose breakdown relative to trehalose synthesis was at least 5-fold more common in the potentially mixotrophic or heterotrophic eukaryotic groups of Alveolata (43%), Amoebozoa (75%), and Excavata (71%) . The other eukaryotic groups

(Archaeplastida, Hacrobia, and Stramenopiles) were more often capable of both breakdown and synthesis (Figure 2 & ~~??~~3). Trehalose has been found to be abundant and to follow a diel cycle intracellularly in *Ostreococcus tauri* (Hirth et al., 2017). Seaweeds, and other marine plants also potentially use trehalose as an osmolyte (Xuan et al., 2012; Danaraj et al., 2020).

3.5 Metatranscriptomic evidence of osmolyte cycling across the surface ocean

In addition to assessing the genetic potential for osmolyte metabolism in marine ~~microbes, we leveraged the Tara prokaryotic and eukaryotic metatranscriptomic data (Salazar et al., 2019; Carradec et al., 2018) to query the~~ microbe monocultures, the metabolism of two key osmolytes, glycine betaine, an amino acid derivative representative, and mannitol, a sugar representative, across the global ocean ~~. We used the~~ was assessed using the Tara prokaryotic and eukaryotic metatranscriptomic data (Salazar et al., 2019; Carradec et al., 2018). We decided to assess the metatranscriptomic signatures of genes associated with targeted osmolytes, as metatranscriptomics provides the ability to assess the transcription and potential activity of these genes across populations (Alexander et al., 2015, 2020; Hu et al., 2018; Salazar et al., 2019). The Ocean Gene Atlas (Villar et al., 2018) was used to assess 1) the diversity of ~~organisms~~ *in situ* communities capable of transport, synthesis, and breakdown of each osmolyte, and 2) the geographical patterns of osmolyte cycling across the surface ocean. Representative ~~KO's~~ KOs were chosen for different subpathways of glycine betaine and mannitol transport, breakdown, and synthesis, and correlations were calculated for the total abundance of orthologs across samples (Supplemental Data Sheet 4). Broadly, we found an alignment in the taxonomic distribution of transcripts recovered with our survey of cultivated genomes and transcriptomes (Figure 1 & 5A). Additionally, we show that transporters were much more highly abundant in the metatranscriptomic datasets compared to either synthesis or breakdown pathways (Figure 5). We also highlight the importance of metazoans in the cycling of glycine betaine (Figure 6).

Subunits for the different transport systems significantly correlated ($p < 0.001$) with each other across samples (Figure 5 A). Proteobacteria dominated both the total expression and the recovered orthologs of both osmolyte transport systems. For mannitol, the substrate-binding subunit (K10227) and permease subunits (K10228, K10229) of the mannitol transport system cluster together and are significantly correlated ($\rho > 0.79$) (Figure 5 A), suggesting that they are co-expressed across the ocean. By contrast, the ATP-binding subunit (K10111, K10112) of the mannitol transport system did not cluster with the other subunits, though one (K10112) was significantly correlated with the associated subunits ($\rho > 0.57$) (Figure 5 A). Two different ABC transporter systems in KEGG are associated with glycine betaine. The glycine betaine/proline transporter is more common than the osmoprotectant transporter (Figure 5 A). All subunits (K02000, K02001, K02002) for the glycine betaine/proline transport system cluster together and are significantly correlated ($\rho > 0.72$). The ATP-binding subunit (K05847) and the substrate-binding subunit (K05845) of the osmoprotectant transporter were significantly correlated ($\rho = 0.59$). The permease subunit (K05846) for the osmoprotectant transporter did not cluster with the other subunits, but was significantly correlated with the ATP-binding and substrate-binding subunits ($\rho > 0.56$) (Figure 5 A).

~~The sugar mannitol~~ Mannitol is often recycled internally, and therefore two mannitol-associated ~~KO's~~ KOs (K00045, K00009) were annotated as being capable of both mannitol breakdown and synthesis (recycling). The recycling of mannitol and fructose (K00045) was highly correlated with three subunits of the mannitol transport system ($\rho > 0.60$). Both the expression and recovered orthologs for K00045 were dominated by Proteobacteria, but also included the greatest representation of Bacteroidetes relative to all other ~~KO's (Figure 5~~ KOs (Figure 5 A). The recycling of mannitol and fructose-6-phosphate (K00009)

was not significantly correlated with any other orthologs, except for K02800 ($\rho = 0.41$) (Figure 5 A). Both the recovered transcripts and expression of mannitol and beta-d-fructose-6-phosphate recycling, as well as the breakdown of mannitol via a phosphotransferase (K02798,K02800), were uniquely dominated by Actinobacteria.

The orthologs for synthesis and breakdown of glycine betaine were overall more abundant than those for mannitol (Figure 5 B). The methyltransferases involved in glycine betaine synthesis from glycine (K18896) and sarcosine (K18897) were significantly correlated ($\rho = 0.88$), and the expression of these orthologs uniquely included a significant portion of Cyanobacteria. Glycine betaine has been previously detected in Cyanobacteria (Fiore et al., 2015; Heal et al., 2020)~~, specifically~~ . Specifically, glycine betaine was characterized in *Synechococcus sp.* WH8102 with the two methyltransferases required for synthesis (Lu et al., 2006). The KO for glycine betaine synthesis from choline (K00108) was the most highly expressed (Figure S5). The expression of orthologs associated with this synthesis pathway (K00499, K00108, K00130) were dominated by Proteobacteria, but were not all significantly correlated with each other. The breakdown of glycine betaine to glycine is a three-step pathway and the associated orthologs were all significantly correlated (K00315, K00302, K00303, K00304) ($\rho > 0.73$). This glycine betaine breakdown cluster was almost exclusively expressed by Proteobacteria and was significantly anti-correlated with the orthologs for glycine betaine transport ($\rho > 0.40$) (Figure 5 A). Interestingly, glycine betaine breakdown was negatively correlated with richness across all samples, suggesting that this pathway is limited to a niche-group, as was observed in the the prokaryotic genomes (Figure S9, Figure 1).

Globally, the mannitol/sorbitol transporter (K10227) was detected in most samples, but expression levels were particularly high towards the poles (Figure 5 B). The latitudinal trend was supported by a significant negative correlation between temperature and mannitol/sorbitol transporter expression ($\rho = -0.66$) (Figure S9). Salinity conditions at the poles are significantly different, and therefore may represent sites of osmolyte recycling that are unique from the majority of the open ocean (Bano et al., 2004). Both mannitol recycling genes (K00045, K00009) were expressed at comparatively low levels (Figure 5 B, Figure S6). Mannitol is generally not considered to be ~~a commonly used osmolyte in~~ commonly used by prokaryotes (Empadinhas and Costa, 2008), but it has been experimentally identified as an osmolyte in several bacteria (Kets et al., 1996; Zahid et al., 2015), though not in any marine bacteria. However, marine bacteria clearly transport mannitol (and/or sorbitol) actively, while expression of the associated cellular recycling genes are orders of magnitude lower (Figure 5 B). While the potential role of mannitol in marine bacteria is not entirely clear, mannitol, like other sugars and polyols, is used as a cryoprotectant in industrial applications where bacterial cells must be preserved through freezing (Savini et al., 2010). We hypothesize that the increased expression of the mannitol/sorbitol transporter in the bacterial size fraction at the poles may be linked to the availability of mannitol through production by eukaryotic phytoplankton, and its cryoprotectant properties for the bacteria.

Both glycine betaine transporters (the glycine betaine/proline transporter, K02002, and the osmoprotectant transporter, K05845) were present in bacterial transcripts. However, the glycine betaine/proline transporter was much more highly expressed throughout the surface ocean and did not show any clear differentiation between high and low latitudes or temperature (Figure 5 B, Figure S9). This expression was predominantly derived from unknown or unclassified bacteria and Proteobacteria. Previous work has shown that the Proteobacteria SAR11 have a high-affinity glycine betaine transporter (Noell and Giovannoni, 2019). *Vibrio parahaemolyticus* species contains two proU ABC transporters, but have also been found to transport glycine betaine, along with dimethylglycine, choline, ectoine, and proline with a BCCT transporter which was not linked to glycine betaine in KEGG (Gregory et al., 2020). Unlike mannitol, a glycine betaine

breakdown gene (K00544) was expressed at an order of magnitude higher than the synthesis gene (K18897), and was, again, predominantly expressed by Proteobacteria ~~Figure 5~~(Figure 5 B, Figure S5). The glycine betaine transport and breakdown expressions suggests that after transport into the cell, glycine betaine may be more commonly recycled through anabolic pathways or catabolized, rather than used as an osmolyte, at least compared to mannitol. Proteobacteria appear to be the most significant prokaryotic group metabolizing glycine betaine in Tara. Nanomolar concentrations of glycine betaine are rapidly scavenged from seawater, and most of this activity is assumed to be driven by bacteria (in the $< 1\mu m$ fraction) (Kiene and Hoffmann Williams, 1998). Radiolabeled glycine betaine was rapidly incorporated in its existing molecular structure, but after 46 hours of incubation, 46% of the radioactivity was transformed into carbon dioxide (Kiene and Hoffmann Williams, 1998), indicating substantial remineralization of glycine betaine as a carbon (and possibly a nitrogen) substrate.

3.6 Zooplankton and the cycling of glycine betaine

Glycine betaine synthesis and breakdown was also present in eukaryotic unigenes (Carradec et al., 2018). The taxonomy of glycine betaine synthesis (K00108, K14085) and breakdown was dominated by Arthropoda (specifically, copepods), but was also found in Vertebrata (fish) and protists (Dinophyceae and Rhizaria) (Figure 6). A similar protistan signature was observed in MMETSP transcriptomes, where glycine betaine breakdown was exclusively observed in the Alveolata (Figure 2). Additionally, genes for glycine betaine synthesis from glycine (K18896, K24071) in Tara were positively correlated with the Dinophyceae pigment marker ($\rho > 0.22$), peridinin, across all samples (Figure S10). As was observed in prokaryotes (Figure 5), glycine betaine synthesis from choline was also most abundant across eukaryotes (Figure S7). However, across all size fractions, total expression by eukaryotes was dominated by glycine betaine breakdown to glycine (Figure 6). The protist contribution to glycine betaine breakdown expression is most likely represented in the $5 - 20\mu m$, whereas the highest expression was contributed by the Arthropoda and Vertebrata in the $180 - 2000\mu m$ fraction.

Copepod excretion is a potentially important source of labile substrates (Maas et al., 2020). Previous studies have described taurine as an important osmolyte in marine copepods (Clifford et al., 2020). Although glycine betaine has been described across all three domains of life (Yancey, 2005), to our knowledge, the potential for glycine betaine synthesis and breakdown has not been described previously for marine planktonic Arthropoda. The expression patterns described here for eukaryotic unigenes hint at an important role for not only protists, but also higher-order marine eukaryotes in the recycling of this osmolyte. The disconnect between breakdown and synthesis across all eukaryotic size fractions suggests that eukaryotes are capable of obtaining glycine betaine from the environment. In the case of ~~the~~ multicellular organisms, glycine betaine could be obtained through grazing or the microbiome (Shoemaker and Moisander, 2017). The dominance of breakdown could in part be due to the conservative threshold used here to filter ortholog hits causing a higher recovery of breakdown ~~KO's~~ KOs and/or removal of synthesis genes. The expression of both synthesis and breakdown of glycine betaine in the $180-2000\mu m$ fraction suggests that multicellular organisms, particularly copepods, could be an important source or sink of this osmolyte that is not currently considered.

4 CONCLUSIONS

Osmolytes are core ~~intracellular~~ cellular metabolites that are required for all marine microbes living in the high salinity environment of the ocean. Across the three domains of life, we found that osmolyte metabolic capabilities could be broken down into two main categories: 1) globally ubiquitous, and 2) mosaically present across groups. The synthesis and breakdown of three amino acid osmolytes surveyed

(glutamate, glutamine, and proline) were common across all prokaryotes and eukaryotes. The ubiquity of synthesis and breakdown of these core metabolites likely indicates an important role for internal recycling. Although it is difficult to directly attribute osmotic function without experimental evidence, we hypothesized that osmolytes in genomes with synthesis only, or genomes with transport only were more likely used for osmotic functions. ~~This trend~~ The genomic evidence for a potential osmotic function (synthesis or transport in the absence of breakdown) was most ~~common~~ apparent for the classically known osmolyte, glycine betaine~~. However, glycine betaine appears~~ , which appeared to be a unique instance of osmolyte specificity in the absence of central metabolic importance. The dominance of synthesis or breakdown of the other osmolytes surveyed differed significantly across prokaryotes and eukaryotes, suggesting that osmolyte sources and sinks are taxa dependent. ~~It is important to note, however, that breakdown was more common than~~ Critically, the ability to breakdown was most common relative to synthesis or transport for the majority of the osmolytes surveyed, suggesting that these small molecules serve a central metabolic function and are likely important substrates in the microbial loop.

We took a two pronged approach in our survey of osmolyte metabolism within marine taxa: 1) examining metabolic potential within reference genomes and transcriptomes and 2) assessing transcriptional activity from mixed communities across the global ocean. When we compared the potential of taxonomic groups to synthesize, breakdown, and transport osmolytes to the expression of the genes associated with these pathways in the environmental metatranscriptomic data, we found good agreement between the species with the potential to perform these metabolic functions, and those that were observed to transcribe the relevant genes in the environment. In the environment, bacterial transcription of transporter genes for both mannitol and glycine betaine were an order of magnitude higher than their respective genes for synthesis and breakdown. Although only four Alphaproteobacteria genomes were found to have the potential for glycine betaine breakdown, the relative abundance of transcription of the breakdown pathway was significantly higher than the breakdown pathway for mannitol. One of the strains capable of breakdown of glycine betaine, Candidatus *Pelagibacter ubique*, is a highly abundant taxa found broadly throughout the oligotrophic ocean (Morris et al., 2002), which perhaps explains the high level of transcription. This suggests that a niche based on glycine betaine as a carbon source may be occupied by these species without competition from other bacterial species. Interestingly, although synthesis of glycine betaine by eukaryotic size fractions was widespread, breakdown was most highly expressed in the larger size fractions, particularly by metazoa. Thus, not only are protists likely important sinks of the osmolyte glycine betaine, but also higher-order, multicellular eukaryotes should also be considered as potential sinks. This finding would not have been possible without probing the Tara metatranscriptomes. Our initial investigation of two key osmolytes within the Tara metatranscriptomes highlights the fact that both eukaryotes and prokaryotes are actively using ~~, recycling, and potentially exchanging osmolytes in the ocean. Taken in concert with our observations of glycine betaine transport and breakdown in the bacterial fraction, this suggests that there is likely much osmolyte exchange between not only phytoplankton, but also zooplanktonwith pelagic bacteria~~and recycling osmolytes *in situ*, and that osmolytes are potentially being exchanged between pelagic bacteria, phytoplankton, and also zooplankton. Marine protists, particularly phytoplankton, are most often emphasized as potential sources of marine osmolytes~~(Durham et al., 2019)~~, but clearly connections with higher order trophic levels are also important in the global cycling of carbon ~~reservoirs~~ and should be ~~more examined in depth (Clifford et al., 2020)~~examined more in depth (Durham et al., 2019; Clifford et al., 2020).

The ocean is a dynamic and variable environment, and the organisms that inhabit it must be prepared to survive sudden environmental fluctuations. Here we surveyed and reviewed the osmoadaptive capabilities ~~,~~ in the form of the synthesis, breakdown, and transport of various osmolytes across cultured organisms with

sequenced genomes or transcriptomes as well as within a global metatranscriptomic dataset. We believe that the work here serves as a first step in the targeted and informed examination of osmoregulation by ecologically-relevant groups in the microbial loop. In particular, further work examining the presence and absence of these genes across metagenome assembled genomes (MAGs) or single celled amplified genomes (SAGs) may provide greater insights into the osmoadaptive capabilities of uncultivated lineages. Additionally, examining rates of uptake and loss of osmolytes from cells, coupled with more comprehensive characterization of their dissolved concentrations in the ocean, would support our understanding of their potentially significant influence on labile carbon cycling in the ocean. Broadly, we show that while some osmolytes (e.g. the amino acids glutamate and glutamine) are ubiquitously synthesized ~~or~~ and broken down across the tree of life, the ability to synthesize, breakdown, and transport most osmolytes is present mosaically across organisms. The mosaicism of the genomic traits of osmolyte metabolism may suggest that many of the osmolytes surveyed here are central currencies of exchange between organisms (Moran et al., 2016), between domains or within domains.

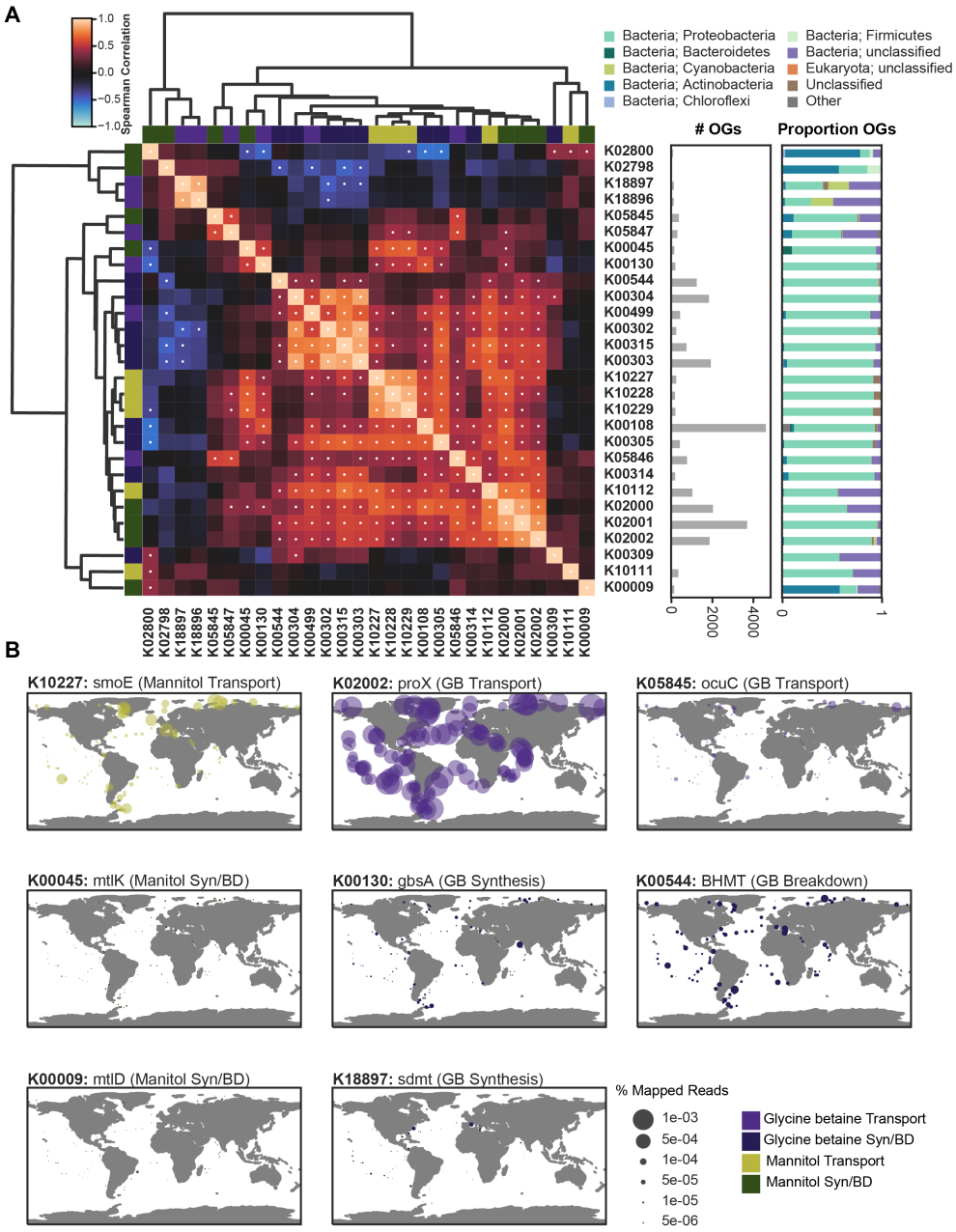

Figure 5. The co-occurrence, taxonomic profile, and global abundance of key orthologs involved in the synthesis, breakdown, and transport of mannitol and glycine betaine. The metatranscriptomic abundance of key orthologs involved in the processing of mannitol (green) and glycine betaine (purple) was assessed across the prokaryotic metatranscriptomic data from Tara Oceans using the OM-RGC_v2 dataset (Figure S6, S5). (A) The relative correlation of metatranscriptomic profiles for each of the orthologs considered was assessed by taking the Spearman's correlation of the sum of all orthologs at a given site. The Spearman's ρ is depicted with a heatmap that is clustered based on Bray-Curtis similarity. Significant correlations were determined with a Bonferroni multi-testing p-value correction, and significant correlations (defined as $p < 0.001$) are depicted as a white dot. The total number of orthologs (OGs) is depicted as a bar graph and the taxonomic breakdown of the OGs is given as a proportion. (B) The relative abundance of key orthologs of interest is plotted as percent mapped reads across all surface samples from the Tara Oceans sampling efforts. Genes involved in mannitol metabolism include: K10111: malK, K10112: msmK, K10227: smoE, K10228: smoF, K10229: smoG, K00009: mtlD, K00045: mtlK, K02800, mtlA, K02798: cmtB. Genes involved in glycine betaine metabolism include: K02000: proV, K02001: proW, K02002: proX, K05845: opuC, K05846: opuBD, K05847: opuA, K00130: gbsA, K00499: CMO, K18896: gsmt, K18897: sdmt, K00108: betA, K00315: DMGDH, K00302: soxA, K00303: soxB, K00304: soxD, K00305: soxG, K00309: dmg, K00544: BHMT, K00314: SARDH.

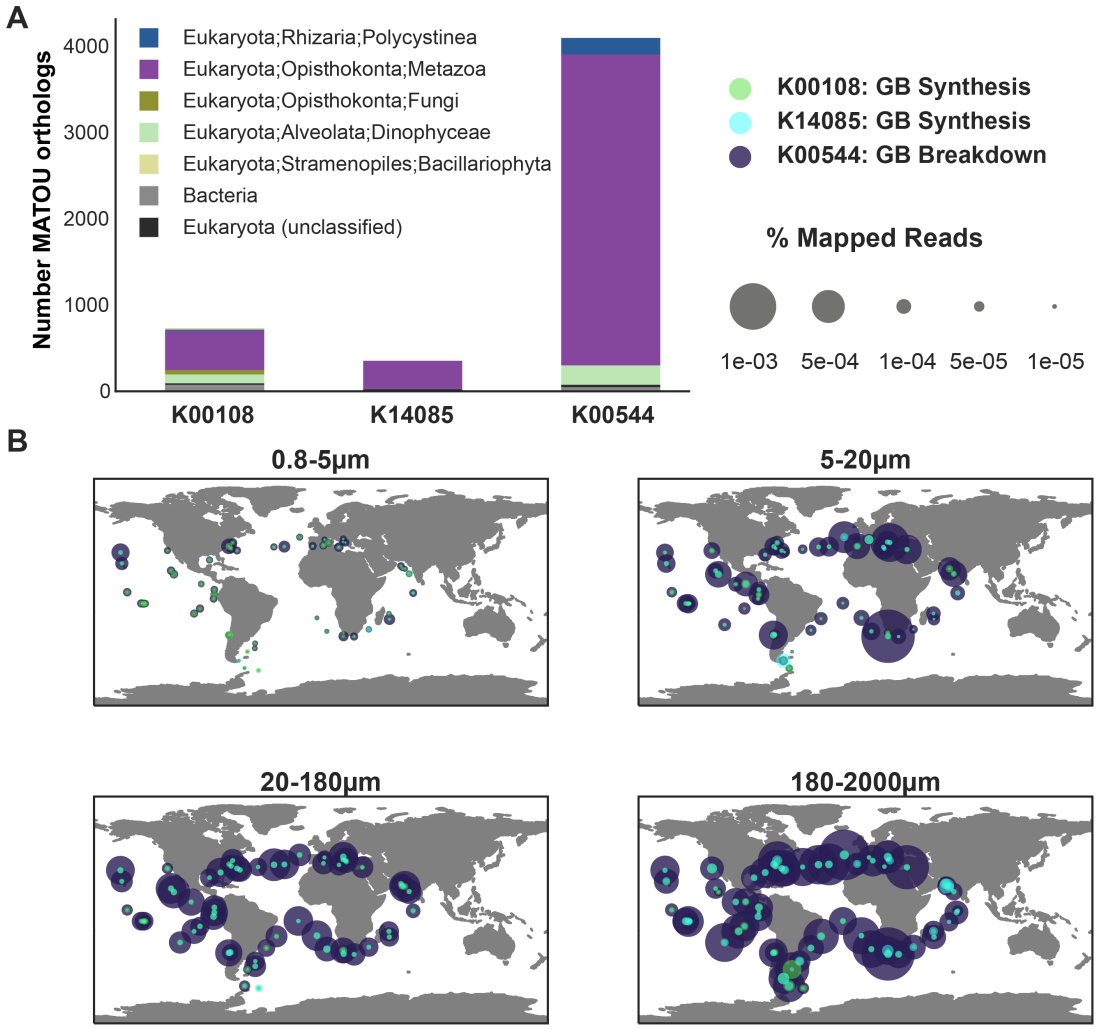

Figure 6. The abundance and diversity of KOs associated with the synthesis and breakdown of glycine betaine across eukaryotic metatranscriptomic data from Tara Oceans. Genes included in the analysis include: Betaine-homocysteine S-methyltransferase (K00544), a central enzyme for the breakdown of glycine betaine, 4267 orthologs of K00544 were identified in the MATOU-v1 dataset; choline dehydrogenase (K00108), 826 orthologs recovered; aldehyde dehydrogenase A1 (K14085), 384 orthologs recovered. (A) The taxonomic breakdown of the orthologs as published in MATOU-v1 is depicted as stacked bar plot (top). (B) The relative abundance of the three KOs (Figure S7, S8) in the eukaryotic metatranscriptome dataset from the surface ocean in Tara is shown by size fraction ($0.8-5\mu m, 5-20\mu m, 20-180\mu m$, and $180-2000\mu m$).

Common name	g/mol	Category	N atoms	S atoms	Cellular	Dissolved (nM)
trigonelline	137	alkaloid	0	1	+	-
TMAO	75	amine oxide	1	0	+	2-77 (Hatton and Gibb, 1999; Gibb and Hatton, 2004)
ectoine	142	amino acid	2	0	+	0.1-0.3 (Widner et al., 2021)
GABA	103	amino acid	1	0	+	0.1 (Widner et al., 2021)
glutamate	147	amino acid	1	0	+	0.7-5 (Widner et al., 2021; Mopper and Lindroth, 1982)
glutamine	146	amino acid	2	0	+	0.2-3 (Widner et al., 2021; Mopper and Lindroth, 1982)
glycine betaine	117	amino acid	1	0	+	-
homarine	137	amino acid	1	0	+	-
homoserine betaine	161	amino acid	1	0	+	-
hydroxyectoine	158	amino acid	2	0	+	-
proline	115	amino acid	1	0	+	0.4 (Widner et al., 2021)
sarcosine	89	amino acid	1	0	+	0.1-0.2 (Widner et al., 2021)
taurine	125	amino acid	1	1	+	0.1-320 (Widner et al., 2021; Clifford et al., 2017)
glucosylglycerol	254	sugar	0	0	+	-
isofloridoside	254	sugar	0	0	+	-
sucrose	342	sugar	0	0	+	BD (Sakugawa and Handa, 1985)
trehalose	342	sugar	0	0	+	BD (Sakugawa and Handa, 1985)
glycerol	92	sugar alcohol	0	0	+	-
mannitol	182	sugar alcohol	0	0	+	-
sorbitol	182	sugar alcohol	0	0	+	-
DHPS	155	sulfonate	0	1	+	0.26-0.44 (Widner et al., 2021)
DMSP	134	sulfonium	0	1	+	1-2 (Kiene and Slezak, 2006)
gonyol	178	sulfonium	0	1	+	-

Table 1. Osmolytes previously described in marine microbes. The table includes: Common name of osmolyte, molecular weight (g/mol), nominal category, number of nitrogen (N) atoms, number of sulfur (S) atoms, cellular concentrations reported (+) in monocultures, and previously measured dissolved concentrations (nM) from coastal or open ocean environments. Names in bold were used in the analyses presented here. As osmolyte cellular concentrations are extremely variable, we only report if an osmolyte has been previously reported in monocultures (Dickson and Kirst, 1987a,b; Brown and Hellebust, 1978; Keller et al., 1989; Pade et al., 2012, 2016; Gebser and Pohnert, 2013; Durham et al., 2019; Fenizia et al., 2020; Yancey, 2005; Mountfort and Pybus, 1992; Reed et al., 1986; Lin et al., 2020). To our knowledge, dissolved measurements of most osmolytes do not exist. Abbreviations: GABA = 4-aminobutanoate, TMAO = trimethylamine n-oxide, DMSP = dimethylsulfoniopropionate, DHPS = dihydroxypropane-sulfonate, BD = below detection

Common name	KEGG id	Breakdown	Synthesis	Transport
ectoine	C06231	2 *(1-3)*	2 *(4-5)*	0
GABA	C00334	~~5~~ 4 *(1-2)*	~~13~~ 11 *(1-4)*	0
glutamate	C00025	~~50~~ 61 *(1-5)*	~~75~~ 78 *(1-3)*	0
glutamine	C00064	30 *(1)*	2 *(1)*	0
glycine betaine	C00719	1 *(3)*	~~6~~ 4 *(1-2)*	2 *(3)*
proline	C00148	~~11~~ 12 *(1)*	~~5~~ 6 *(1-2)*	0
sarcosine	C00213	~~4~~ 2 *(1)*	~~9~~ 7 *(1-4)*	0
taurine	C00245	6 *(1)*	~~4~~ 5 *(1)*	1 *(3)*
glycerol	C00116	~~8~~ 7 *(1-3)*	~~11~~ 9 *(1-4)*	1 *(3)*
mannitol	C00392	~~4~~ 2 *(1-2)*	~~3~~ 1 *(2)*	1 *(3)*
sorbitol	C00794	~~5~~ 4 *(1)*	~~5~~ 4 *(1)*	1 *(3)*
sucrose	C00089	~~13~~ 12 *(1)*	~~2~~ 3 *(1-3)*	0
trehalose	C01083	6 *(1)*	~~4~~ 5 *(1-3)*	0
TMAO	C01104	1 *(1)*	1 *(1)*	0

Table 2. Osmolytes included in prokaryote and eukaryote surveys. The table includes: Common name of osmolyte, the KEGG ID, ~~and~~ the total number of curated pathways with associated KOs used in the searches for ~~Breakdown~~breakdown, ~~Synthesis~~synthesis, and ~~Transport. In some cases~~transport, ~~one~~ and the range of steps in each pathway ~~may represent~~ in parentheses. Some pathways required the presence of multiple steps (i.e. KOs) to be considered complete (Supplemental Data Sheet 1). A single step value indicates all pathways had the same number of steps. For osmolyte pathways with differing number of steps, a range is reported which reflects the minimum and maximum number of steps for all pathways. Abbreviations: GABA = 4-aminobutanoate, TMAO = trimethylamine n-oxide

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

All three authors contributed equally to the development of this project, analysis of the data, and writing of the manuscript.

FUNDING

ELM was supported by the Postdoctoral Scholar Program at Woods Hole Oceanographic Institution. WMJ was supported by a Research Initiative Award from the College of Arts and Sciences at the University of North Carolina Wilmington. HA was supported by a Independent Research and Development Award from the Woods Hole Oceanographic Institution.

ACKNOWLEDGMENTS

This research would not have been possible without the community driven efforts to provide open and freely available data by the Tara Oceans Consortium, the Marine Microbial Transcriptome Sequencing Project (MMETSP), the Marine Metagenomics Portal, and the Kyoto Encyclopedia of Genes and Genomes (KEGG).

DATA AVAILABILITY STATEMENT

These analyses were based on previously existing datasets including the MMETSP (https://zenodo.org/record/1212585), Tara Oceans Gene Atlas (https://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/), KEGG homologs (https://www.genome.jp/tools/kofamkoala/), and MarRef v5 (https://mmp.sfb.uit.no/downloads/).

REFERENCES

Alexander, H., Rouco, M., Haley, S. T., and Dyhrman, S. T. (2020). Transcriptional response of *Emiliania huxleyi* under changing nutrient environments in the North Pacific Subtropical Gyre. *Environmental Microbiology* , 1462–2920.14942doi:10.1111/1462-2920.14942

Alexander, H., Rouco, M., Haley, S. T., Wilson, S. T., Karl, D. M., and Dyhrman, S. T. (2015). Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proceedings of the National Academy of Sciences* 112, E5972–E5979. doi:10.1073/pnas.1518165112

Andreae, M. O. (1986). The ocean as a source of atmospheric sulfur compounds. In *The Role of Air-Sea Exchange in Geochemical Cycling* (Springer Netherlands). 331–362. doi:10.1007/978-94-009-4738-2_14

Arakawa, T. and Timasheff, S. N. (1985). The stabilization of proteins by osmolytes. *Biophysical Journal* 47, 411–414. doi:10.1016/S0006-3495(85)83932-1

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2019). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi:10.1093/bioinformatics/btz859

Bano, N., Ruffin, S., Ransom, B., and Hollibaugh, J. T. (2004). Phylogenetic composition of arctic ocean archaeal assemblages and comparison with antarctic assemblages. *Applied and Environmental Microbiology* 70, 781–789. doi:10.1128/AEM.70.2.781-789.2004

Bisson, M. and Kirst, G. (1995). Osmotic acclimation and turgor pressure regulation in algae. *Naturwissenschaften* 82, 461–471

Boyd, C. M. and Gradmann, D. (2002). Impact of osmolytes on buoyancy of marine phytoplankton. *Marine Biology* 141, 605–618. doi:10.1007/s00227-002-0872-z

Boysen, A. K., Carlson, L. T., Durham, B. P., Groussman, R. D., Aylward, F. O., Ribalet, F., et al. (2020). Diel oscillations of particulate metabolites reflect synchronized microbial activity in the North Pacific subtropical gyre. *bioRxiv* , 2020.05.09.086173doi:10.1101/2020.05.09.086173

Brand, L. E. (1984). The Salinity Phytoplankton Tolerance Isolates of Forty-six Marine phytoplankton species. *Estuarine, Coastal and Shelf Science* 18, 543–556

Brands, S., Schein, P., Castro-Ochoa, K. F., and Galinski, E. A. (2019). Hydroxyl radical scavenging of the compatible solute ectoine generates two N-acetimides. *Archives of Biochemistry and Biophysics* 674. doi:10.1016/j.abb.2019.108097

Brill, J., Hoffmann, T., Bleisteiner, M., and Bremer, E. (2011). Osmotically controlled synthesis of the compatible solute proline is critical for cellular defense of Bacillus subtilis against high osmolarity. *Journal of Bacteriology* 193, 5335–5346. doi:10.1128/JB.05490-11

Brown, L. M. and Hellebust, J. A. (1978). Sorbitol and proline as intracellular osmotic solutes in the green alga stichococcus bacillaris. *Canadian Journal of Botany* 56, 676–679. doi:10.1139/b78-074

Burg, M. B. and Ferraris, J. D. (2008). Intracellular organic osmolytes: Function and regulation. *Journal of Biological Chemistry* 283, 7309–7313. doi:10.1074/jbc.R700042200

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018). A global ocean atlas of eukaryotic genes. *Nature Communications* 9, 373. doi:10.1038/

s41467-017-02342-1
Challenger, F. (1951). Biological methylation. *Advances in Enzymology - and Related Areas of Molecular Biology* 12, 429–491
Clifford, E. L., De Corte, D., Amano, C., Paliaga, P., Ivančić, I., Ortiz, V., et al. (2020). Mesozooplankton taurine production and prokaryotic uptake in the northern Adriatic Sea. *Limnology and Oceanography* 1, 1–18. doi:10.1002/lno.11544
Clifford, E. L., Hansell, D. A., Varela, M. M., Nieto-Cid, M., Herndl, G. J., and Sintes, E. (2017). Crustacean zooplankton release copious amounts of dissolved organic matter as taurine in the ocean. *Limnology and Oceanography* 62, 2745–2758. doi:10.1002/lno.10603
Clifford, E. L., Varela, M. M., De Corte, D., Bode, A., Ortiz, V., Herndl, G., et al. (2019). Taurine is a major carbon and energy source for marine prokaryotes in the North Atlantic Ocean off the Iberian Peninsula. *Microbiology of Aquatic Systems* doi:10.1071/MU9760120
Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi:10.1093/bioinformatics/btp163
Danaraj, J., Yosuva, M., and Ayyappan, S. (2020). Comparative metabolomics analysis of wild and suspension cultured cells (SCC) of seagrass Halodule pinifolia (Miki) hartog of cymodoceaceae family. *Aquatic Botany* 167, 103278. doi:10.1016/j.aquabot.2020.103278
Dawson, H. M., Heal, K. R., Boysen, A. K., Carlson, L. T., Ingalls, A. E., and Young, J. N. (2020a). Potential of temperature- and salinity-driven shifts in diatom compatible solute concentrations to impact biogeochemical cycling within sea ice. *Elementa* 8. doi:10.1525/elementa.421
Dawson, H. M., Heal, K. R., Torstensson, A., Carlson, L. T., Ingalls, A. E., and Young, J. N. (2020b). Large Diversity in Nitrogen- and Sulfur-Containing Compatible Solute Profiles in Polar and Temperate Diatoms. *Integrative and Comparative Biology* 60, 1401–1413. doi:10.1093/icb/icaa133
Dickson, D. M. J. and Kirst, G. O. (1987a). Osmotic adjustment in marine eukaryotic algae: The role of inorganic ions, quaternary ammonium, tertiary sulphonium and carbohydrate solutes. I. Diatoms and a rhodophyte. *The New Phytologist* 106, 645–655
Dickson, D. M. J. and Kirst, G. O. (1987b). Osmotic adjustment in marine eukaryotic algae: The role of inorganic ions, quaternary ammonium, tertiary sulphonium and carbohydrate solutes. II. Prasinophytes and haptophytes. *The New Phytologist* 106, 645–655
Ducklow, H. (2000). Bacterial production and biomass in the oceans. In *Microbial Ecology of the Oceans*, ed. D. L. Kirchman (New York: Wiley), chap. 4. 85–120
Durham, B. P., Boysen, A. K., Carlson, L. T., Groussman, R. D., Heal, K. R., Cain, K. R., et al. (2019). Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nature Microbiology* 4, 1706–1715. doi:10.1038/s41564-019-0507-5
Empadinhas, N. and Costa, M. S. (2008). Osmoadaptation mechanisms in prokaryotes : distribution of compatible solutes. *International Microbiology* 11, 151–161. doi:10.2436/20.1501.01.55
Fenizia, S., Thume, K., Wirgenings, M., and Pohnert, G. (2020). Ectoine from bacterial and algal origin is a compatible solute in microalgae. *Marine Drugs* 18, 42. doi:10.3390/md18010042
Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281, 237–240. doi:10.1126/science.281.5374.237
Fiore, C. L., Longnecker, K., Kido Soule, M. C., and Kujawinski, E. B. (2015). Release of Ecologically Relevant Metabolites by the Cyanobacterium, Synechococcus elongatus CCMP 1631. *Environmental Microbiology* 17, n/a–n/a. doi:10.1111/1462-2920.12899

838 Gao, Y., Xi, Y., Lu, X. L., Zheng, H., Hu, B., Liu, X. Y., et al. (2013). Cloning, expression and functional
839 characterization of a novel trehalose synthase from marine Pseudomonas sp. P8005. *World Journal of*
840 *Microbiology and Biotechnology* 29, 2195–2206. doi:10.1007/s11274-013-1385-2
841 Gebser, B. and Pohnert, G. (2013). Synchronized regulation of different zwitterionic metabolites in the
842 osmoadaption of phytoplankton. *Marine drugs* 11, 2168–2182. doi:10.3390/md11062168
843 Gibb, S. W. and Hatton, A. D. (2004). The occurrence and distribution of trimethylamine-N-oxide in
844 Antarctic coastal waters. *Marine Chemistry* 91, 65–75. doi:10.1016/j.marchem.2004.04.005
845 Goode, K. L., Dunphy, B. J., and Parsons, D. M. (2020). Environmental metabolomics as an ecological
846 indicator: Metabolite profiles in juvenile fish discriminate sites with different nursery habitat qualities.
847 *Ecological Indicators* 115. doi:10.1016/j.ecolind.2020.106361
848 Gregory, G. J., Dutta, A., Parashar, V., and Boyd, E. F. (2020). Investigations of dimethylglycine, glycine
849 betaine, and ectoine uptake by a betaine-carnitine-choline transporter family transporter with diverse
850 substrate specificity in vibrio species. *Journal of Bacteriology* 202. doi:10.1128/JB.00314-20
851 Götz, F., Longnecker, K., Soule, M. C. K., Becker, K. W., McNichol, J., Kujawinski, E. B., et al. (2018).
852 Targeted metabolomics reveals proline as a major osmolyte in the chemolithoautotroph sulfurimonas
853 denitrificans. *MicrobiologyOpen* 7. doi:10.1002/mbo3.586
854 Hatton, A. D. and Gibb, S. W. (1999). A technique for the determination of trimethylamine-N-oxide in
855 natural waters and biological media. *Analytical Chemistry* 71, 4886–4891. doi:10.1021/ac990366y
856 Heal, K. R., Durham, B., Boysen, A. K., Carlson, L. T., Qin, W., Ribalet, F., et al. (2020). Marine
857 community metabolomes carry fingerprints of phytoplankton community composition. *bioRxiv* doi:10.
858 1101/2020.12.22.424086
859 Hirth, M., Liverani, S., Mahlow, S., Bouget, F. Y., Pohnert, G., and Sasso, S. (2017). Metabolic profiling
860 identifies trehalose as an abundant and diurnally fluctuating metabolite in the microalga Ostreococcus
861 tauri. *Metabolomics* 13, 0. doi:10.1007/s11306-017-1203-1
862 Hosoi, M., Takeuchi, K., Sawada, H., and Toyohara, H. (2005). Expression and functional analysis of
863 mussel taurine transporter, as a key molecule in cellular osmoconforming. *The Journal of experimental*
864 *biology* 208, 4203–11. doi:10.1242/jeb.01868
865 Hu, S. K., Liu, Z., Alexander, H., Campbell, V., Connell, P. E., Dyhrman, S. T., et al. (2018). Shifting
866 metabolic priorities among key protistan taxa within and below the euphotic zone. *Environmental*
867 *Microbiology* 20, 2865–2879. doi:10.1111/1462-2920.14259
868 Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A
869 new view of the tree of life. *Nat Microbiol* 1. doi:10.1038/nmicrobiol.2016.48
870 Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* 9,
871 90–95. doi:10.1109/MCSE.2007.55
872 Johnson, L. K., Alexander, H., and Brown, C. T. (2018). Re-assembly, quality evaluation, and annotation
873 of 678 microbial eukaryotic reference transcriptomes. *GigaScience* 8. doi:10.1093/gigascience/giy158
874 Johnson, W. M., Kido Soule, M. C., Longnecker, K., Bhatia, M. P., Hallam, S. J., Lomas, M. W., et al.
875 (2021). Insights into the controls on metabolite distributions along a latitudinal transect of the western
876 atlantic ocean. *bioRxiv* doi:10.1101/2021.03.09.434501
877 Johnson, W. M., Longnecker, K., Kido Soule, M. C., Arnold, W. A., Bhatia, M. P., Hallam, S. J., et al.
878 (2020). Metabolite composition of sinking particles differs from surface suspended particles across a
879 latitudinal transect in the South Atlantic. *Limnology and Oceanography* 65, 111–127. doi:10.1002/lno.
880 11255
881 Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids*
882 *research* 28, 27–30

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research* 40, D109–D114. doi:10.1093/nar/gkr988

Kaspar, H. F., Mountfort, D. O., and Pybus, V. (1991). Degradation of gamma-aminobutyric acid (GABA) by marine microorganisms. *FEMS Microbiology Letters* 85, 313–318. doi:10.1111/j.1574-6968.1991.tb04757.x

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., et al. (2014). The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biology* 12, e1001889. doi:10.1371/journal.pbio.1001889

Keller, M., Bellows, W., and Guillard, R. (1989). Dimethyl Sulfide production in marine phytoplankton. In *Biogenic sulfur in the environment*, eds. E. Saltzman and W. Cooper (Washington D.C.: American Chemical Society). 167–182

Keller, M. D., Kiene, R. P., Matrai, P. A., and Bellows, W. K. (1999a). Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. I. Batch cultures. *Marine Biology* 135, 237–248. doi:10.1007/s002270050621

Keller, M. D., Kiene, R. P., Matrai, P. A., and Bellows, W. K. (1999b). Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. II. N-limited chemostat cultures. *Marine Biology* 135, 249–257. doi:10.1007/s002270050622

Kets, E. P., Galinski, E. A., De Wit, M., De Bont, J. A., and Heipieper, H. J. (1996). Mannitol, a novel bacterial compatible solute in Pseudomonas putida S12. *Journal of Bacteriology* 178, 6665–6670. doi:10.1128/jb.178.23.6665-6670.1996

Kiene, R. P. and Hoffmann Williams, L. P. (1998). Glycine betaine uptake, retention, and degradation by microorganisms in seawater. *Limnology and Oceanography* 43, 1592–1603. doi:10.4319/lo.1998.43.7.1592

Kiene, R. P. and Linn, L. J. (2000). Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico. *Limnology and Oceanography* 45, 849–861. doi:10.4319/lo.2000.45.4.0849

Kiene, R. P. and Slezak, D. (2006). Low dissolved DMSP concentrations in seawater revealed by small volume gravity filtration and dialysis sampling. *Limnology and Oceanography: Methods* 4, 80–95. doi:10.4319/lom.2006.4.80

Kinne, R. K. (1993). The role of organic osmolytes in osmoregulation: From bacteria to mammals. *Journal of Experimental Zoology* 265, 346–355. doi:10.1002/jez.1402650403

Kirst, G. O. (1989). Salinity Tolerance of Eukaryotic Marine Algae. *Annual Review of Plant Physiology and Plant Molecular Biology* 40, 21–53. doi:10.1146/annurev.pp.41.060190.000321

Klähn, S. and Hagemann, M. (2011). Compatible solute biosynthesis in cyanobacteria. *Environmental Microbiology* 13, 551–562. doi:10.1111/j.1462-2920.2010.02366.x

Klemetsen, T., Raknes, I. A., Fu, J., Agafonov, A., Balasundaram, S. V., Tartari, G., et al. (2017). The MAR databases: development and implementation of databases specific for marine metagenomics. *Nucleic Acids Research* 46, D692–D699. doi:10.1093/nar/gkx1036

Lavoie, M., Levasseur, M., and Babin, M. (2015). Testing the potential ballast role for dimethylsulfoniopropionate in marine phytoplankton: A modeling study. *Journal of Plankton Research* 37, 699–711. doi:10.1093/plankt/fbv050

Lavoie, M., Waller, J. C., Kiene, R. P., and Levasseur, M. (2018). Polar marine diatoms likely take up a small fraction of dissolved dimethylsulfoniopropionate relative to bacteria in oligotrophic environments.

928 *Aquatic Microbial Ecology* 81, 213–218
929 Lee, M. D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, 4162–4164.
930 doi:10.1093/bioinformatics/btz188
931 Letunic, I. and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and
932 annotation of phylogenetic and other trees. *Nucleic Acids Res* 44, W242–W245. doi:10.1093/nar/gkw290
933 Lin, P.-C., Zhang, F., and Pakrasi, H. B. (2020). Enhanced production of sucrose in the fast-growing
934 cyanobacterium Synechococcus elongatus UTEX 2973. *Scientific Reports* 10, 390. doi:10.1038/
935 s41598-019-57319-5
936 Lu, W. D., Chi, Z. M., and Su, C. D. (2006). Identification of glycine betaine as compatible solute in
937 synechococcus sp. wh8102 and characterization of its n-methyltransferase genes involved in betaine
938 synthesis. *Archives of Microbiology* 186, 495–506. doi:10.1007/s00203-006-0167-8
939 Ma, Y., Wang, Q., Gao, X., and Zhang, Y. (2017). Biosynthesis and uptake of glycine betaine as cold-stress
940 response to low temperature in fish pathogen Vibrio anguillarum. *Journal of Microbiology* 55, 44–55.
941 doi:10.1007/s12275-017-6370-2
942 Maas, A. E., Liu, S., Bolaños, L. M., Widner, B., Parsons, R., Kujawinski, E. B., et al. (2020). Migratory
943 zooplankton excreta and its influence on prokaryotic communities. *Frontiers in Marine Science* 7, 1014.
944 doi:10.3389/fmars.2020.573268
945 Masuda, K., Guo, X.-f., Uryu, N., Hagiwara, T., and Watabe, S. (2008). Isolation of marine yeasts
946 collected from the Pacific Ocean showing a high production of gamma-aminobutyric acid. *Bioscience,*
947 *biotechnology, and biochemistry* 72, 3265–3272. doi:10.1271/bbb.80544
948 Mayfield, A. B. and Gates, R. D. (2007). Osmoregulation in anthozoan–dinoflagellate symbiosis.
949 *Comparative Biochemistry and Physiology* 147, 1–10. doi:10.1016/j.cbpa.2006.12.042
950 McParland, E. L. and Levine, N. M. (2019). The role of differential DMSP production and community
951 composition in predicting variability of global surface DMSP concentrations. *Limnology and*
952 *Oceanography* , 1–17doi:10.1002/lno.11076
953 McParland, E. L., Wright, A., Art, K., He, M., and Levine, N. M. (2020). Evidence for contrasting
954 roles of dimethylsulfoniopropionate production in Emiliania huxleyi and Thalassiosira oceanica. *New*
955 *Phytologist* 226, 396–409
956 Mee, M. T. and Wang, H. H. (2012). Engineering ecosystems and synthetic ecologies. *Molecular*
957 *BioSystems* 8, 2470–2483. doi:10.1039/c2mb25133g
958 Met Office (2010 - 2015). *Cartopy: a cartographic python library with a matplotlib interface*. Exeter,
959 Devon
960 Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search:
961 HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research* 41, e121–e121.
962 doi:10.1093/nar/gkt263
963 Mopper, K. and Lindroth, P. (1982). Diel and depth variations in dissolved free amino acids and ammonium
964 in the Baltic Sea determined by shipboard HPLC analysis. *Limnology and Oceanography* 27, 336–347.
965 doi:10.4319/lo.1982.27.2.0336
966 Moran, M. A., Kujawinski, E. B., Stubbins, A., Fatland, R., Aluwihare, L. I., Buchan, A., et al. (2016).
967 Deciphering ocean carbon in a changing world. *Proceedings of the National Academy of Sciences of the*
968 *United States of America* 113, 3143–3151. doi:10.1073/pnas.1514645113
969 Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., et al.
970 (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420, 806–810.
971 doi:10.1038/nature01240

Motard-Côté, J. and Kiene, R. P. (2015). Osmoprotective role of dimethylsulfoniopropionate (DMSP) for estuarine bacterioplankton. *Aquatic Microbial Ecology* 76, 133–147. doi:10.3354/ame01772

Mountfort, D. O. and Pybus, V. (1992). Effect of ph, temperature and salinity on the production of gamma aminobutyric acid (GABA) from amines by marine bacteria. *FEMS Microbiology Letters* 101, 237–244. doi:10.1111/j.1574-6968.1992.tb05780.x

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., et al. (2021). Sustainable data analysis with snakemake. *F1000Res* 10, 33. doi:10.12688/f1000research.29032.1

Nelson, K. E., Eisen, J. A., and Fraser, C. M. (2001). Genome of thermotoga maritima msb8. In *Hyperthermophilic Enzymes Part A* (Academic Press), vol. 330 of *Methods in Enzymology*. 169–180. doi:https://doi.org/10.1016/S0076-6879(01)30374-9

Noell, S. E. and Giovannoni, S. J. (2019). SAR11 bacteria have a high affinity and multifunctional glycine betaine transporter. *Environmental Microbiology* 21, 2559–2575. doi:10.1111/1462-2920.14649

Ongagna-Yhombi, S. Y. and Boyd, F. E. (2013). Biosynthesis of the osmoprotectant ectoine, but not glycine betaine, is critical for survival of osmotically stressed Vibrio parahaemolyticus cells. *Applied and Environmental Microbiology* 79, 5038–5049. doi:10.1128/AEM.01008-13

Ono, H., Sawada, K., Khunajakr, N., Tao, T., Yamamoto, M., Hiramoto, M., et al. (1999). Characterization of biosynthetic enzymes for ectoine as a compatible solute in a moderately halophilic eubacterium, Halomonas elongata. *Journal of Bacteriology* 181, 91–99. doi:10.1128/jb.181.1.91-99.1999

Pachiadaki, M. G., Brown, J. M., Brown, J., Bezuidt, O., Berube, P. M., Biller, S. J., et al. (2019). Charting the complexity of the marine microbiome through single-cell genomics. *Cell* 179, 1623–1635.e11. doi:10.1016/j.cell.2019.11.017

Pade, N., Compaoré, J., Klähn, S., Stal, L. J., and Hagemann, M. (2012). The marine cyanobacterium Crocosphaera watsonii WH8501 synthesizes the compatible solute trehalose by a laterally acquired OtsAB fusion protein. *Environmental Microbiology* 14, 1261–1271. doi:10.1111/j.1462-2920.2012.02709.x

Pade, N., Michalik, D., Ruth, W., Belkin, N., Hess, W. R., Berman-Frank, I., et al. (2016). Trimethylated homoserine functions as the major compatible solute in the globally significant oceanic cyanobacterium Trichodesmium. *Proceedings of the National Academy of Sciences of the United States of America* 113, 13191–13196. doi:10.1073/pnas.1611666113

[Dataset] pandas development team, T. (2020). pandas-dev/pandas: Pandas. doi:10.5281/zenodo.3509134

Pflughoeft, K. J., Kierek, K., and Watnick, P. I. (2003). Role of Ectoine in Vibrio cholerae Osmoadaptation. *Applied and Environmental Microbiology* 69, 5919–5927. doi:10.1128/AEM.69.10.5919

Poddar, N., Sen, R., and Martin, G. J. (2018). Glycerol and nitrate utilisation by marine microalgae nannochloropsis salina and chlorella sp. and associated bacteria during mixotrophic and heterotrophic growth. *Algal Research* 33, 298–309. doi:10.1016/j.algal.2018.06.002

Poli, A., Finore, I., Romano, I., Gioiello, A., Lama, L., and Nicolaus, B. (2017). Microbial Diversity in Extreme Marine Habitats and Their Biomolecules. *Microorganisms* 5, 25. doi:10.3390/microorganisms5020025

Poretsky, R. S., Sun, S., Mou, X., and Moran, M. A. (2010). Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environmental Microbiology* 12, 616–627. doi:10.1111/j.1462-2920.2009.02102.x

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2: Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5, e9490. doi:10.1371/journal.pone.0009490

Reed, R., Borowitzka, L., Mackay, M., Chudek, J., Foster, R., Warr, S., et al. (1986). Organic solute accumulation in osmotically stressed cyanobacteria. *FEMS Microbiology Letters* 39, 51–56. doi:https:

//doi.org/10.1111/j.1574-6968.1986.tb01842.x
Reitzer, L. (2003). Nitrogen assimilation and global regulation in Escherichia coli. *Annual Review of Microbiology* 57, 155–176. doi:10.1146/annurev.micro.57.030502.090820
Roeßler, M. and Muller, V. (2001). Osmoadaptation in bacteria and archaea: common principles and differences. *Environmental Microbiology* 3, 743–754. doi:10.1046/j.1462-2920.2001.00252.x
Sakugawa, H. and Handa, N. (1985). Chemical studies on dissolved carbohydrates in the water samples collected from the north pacific and bering sea. *OCEANOLOGICA ACTA* 8, 185–196
Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Cuenca, M., et al. (2019). Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* 179, 1068–1083.e21. doi:10.1016/j.cell.2019.10.014
Sand, M., Mingote, A. I., Santos, H., Müller, V., and Averhoff, B. (2013). Mannitol, a compatible solute synthesized by acinetobacter baylyi in a two-step pathway including a salt-induced and salt-dependent mannitol-1-phosphate dehydrogenase. *Environmental Microbiology* doi:10.1111/1462-2920.12090
Saum, S. H. and Müller, V. (2008). Regulation of osmoadaptation in the moderate halophile Halobacillus halophilus: Chloride, glutamate and switching osmolyte strategies. *Saline Systems* 4, 1–15. doi:10.1186/1746-1448-4-4
Savini, M., Cecchini, C., Verdenelli, M. C., Silvi, S., Orpianesi, C., and Cresci, A. (2010). Pilot-scale production and viability analysis of freeze-dried probiotic bacteria using different protective agents. *Nutrients* 2, 330–339. doi:10.3390/nu2030330
Schulz, A., Stöveken, N., Binzen, I. M., Hoffmann, T., Heider, J., and Bremer, E. (2017). Feeding on compatible solutes: A substrate-induced pathway for uptake and catabolism of ectoines and its genetic control by EnuR. *Environmental Microbiology* 19, 926–946. doi:10.1111/1462-2920.13414
Shoemaker, K. M. and Moisander, P. H. (2017). Seasonal variation in the copepod gut microbiome in the subtropical north atlantic ocean. *Environmental Microbiology* 19, 3087–3097. doi:https://doi.org/10.1111/1462-2920.13780
Smirnoff, N. and Cumbes, Q. J. (1989). Hydroxyl radical scavenging activity of compatible solutes. *Phytochemistry* 28, 1057–1060. doi:10.1016/0031-9422(89)80182-7
Spielmeyer, A., Gebser, B., and Pohnert, G. (2011). Investigations of the uptake of dimethylsulfoniopropionate by phytoplankton. *ChemBioChem* 12, 2276–2279. doi:10.1002/cbic.201100416
Stefels, J. (2000). Physiological aspects of the production and conversion of DMSP in marine algae and higher plants. *Journal of Sea Research* 43, 183–197
Strom, A. R. (1979). Biosynthesis of trimethylamine oxide in calanoid copepods. seasonal changes in trimethylamine monooxygenase activity. *Marine Biology* 51, 33–40. doi:10.1007/BF00389028
Suescún-Bolívar, L. P., Traverse, G. M. I., and Thomé, P. E. (2016). Glycerol outflow in symbiodinium under osmotic and nitrogen stress. *Mar Biol* 163. doi:10.1007/s00227-016-2899-6
Sunda, W., Kieber, D. J., Kiene, R. P., and Huntsman, S. (2002). An antioxidant function for DMSP and DMS in marine algae. *Nature* 418, 317–320
Takasu, H. and Nagata, T. (2015). High proline content of bacteria-sized particles in the Western North Pacific and its potential as a new biogeochemical indicator of organic matter diagenesis. *Frontiers in Marine Science* 2, 110. doi:10.3389/fmars.2015.00110
ten Hoopen, P., Finn, R. D., Bongo, L. A., Corre, E., Fosso, B., Meyer, F., et al. (2017). The metagenomic data life-cycle: standards and best practices. *GigaScience* 6. doi:10.1093/gigascience/gix047
Treberg, J. R., Speers-Roesch, B., Piermarini, P. M., Ip, Y. K., Ballantyne, J. S., and Driedzic, W. R. (2006). The accumulation of methylamine counteracting solutes in elasmobranchs with differing levels

of urea: A comparison of marine and freshwater species. *Journal of Experimental Biology* 209, 860–870. doi:10.1242/jeb.02055

Villar, E., Vannier, T., Vernette, C., Lescot, M., Cuenca, M., Alexandre, A., et al. (2018). The ocean gene atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Research* 46, W289–W295. doi:10.1093/nar/gky376

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2

Vorobev, A., Sharma, S., Yu, M., Lee, J., Washington, B. J., Whitman, W. B., et al. (2018). Identifying labile dom components in a coastal ocean through depleted bacterial transcripts and chemical signals. *Environmental Microbiology* 20, 1–19. doi:10.1111/1462-2920.14344

[Dataset] Waskom, M. and the seaborn development team (2020). mwaskom/seaborn. doi:10.5281/zenodo.592845

Widner, B., Kido Soule, M. C., Ferrer-González, F. X., Moran, M. A., and Kujawinski, E. B. (2021). Quantification of Amine- and Alcohol-Containing Metabolites in Saline Samples Using Pre-extraction Benzoyl Chloride Derivatization and Ultrahigh Performance Liquid Chromatography Tandem Mass Spectrometry (UHPLC MS/MS). *Analytical Chemistry* 93, 4809–4817. doi:10.1021/acs.analchem.0c03769

Williams, T. J., Long, E., Evans, F., Demaere, M. Z., Lauro, F. M., Raftery, M. J., et al. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME Journal* 6, 1883–1900. doi:10.1038/ismej.2012.28

Xuan, J., Feng, Y., Weng, M., Zhao, G., Shi, J., Yao, J., et al. (2012). Expressed sequence tag analysis and cloning of trehalose-6-phosphate synthase gene from marine alga Laminaria japonica (Phaeophyta). *Acta Oceanologica Sinica* 31, 139–148. doi:10.1007/s13131-012-0260-6

Yancey, P. H. (2005). Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *The Journal of experimental biology* 208, 2819–2830. doi:10.1242/jeb.01730

Yancey, P. H., Blake, W. R., and Conley, J. (2002). Unusual organic osmolytes in deep-sea animals: Adaptations to hydrostatic pressure and other perturbants. *Comparative Biochemistry and Physiology Part A* 133, 667–676. doi:10.1016/S1095-6433(02)00182-4

Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D., and Somero, G. N. (1982). Classes of intracellular osmolyte systems and their distributions living with water stress: Evolution of osmolyte systems. *Science* 217, 1214–1222. doi:10.1126/science.7112124

Yancey, P. H. and Siebenaller, J. F. (2015). Co-evolution of proteins and solutions: Protein adaptation versus cytoprotective micromolecules and their roles in marine organisms. *Journal of Experimental Biology* 218, 1880–1896. doi:10.1242/jeb.114355

Yin, M., Palmer, H. R., Fyfe-Johnson, A. I., Bedford, J. J., Smith, R. A. J., and Yancey, P. H. (2000). Hypotaurine, N-methyltaurine, taurine, and glycine betaine as dominant osmolytes of vestimentiferan tubeworms from hydrothermal vents and cold seeps. *Physiological and Biochemical Zoology* 73, 629–637

Zahid, N., Schweiger, P., Galinski, E., and Deppenmeier, U. (2015). Identification of mannitol as compatible solute in Gluconobacter oxydans. *Applied Microbiology and Biotechnology* 99, 5511–5521. doi:10.1007/s00253-015-6626-x