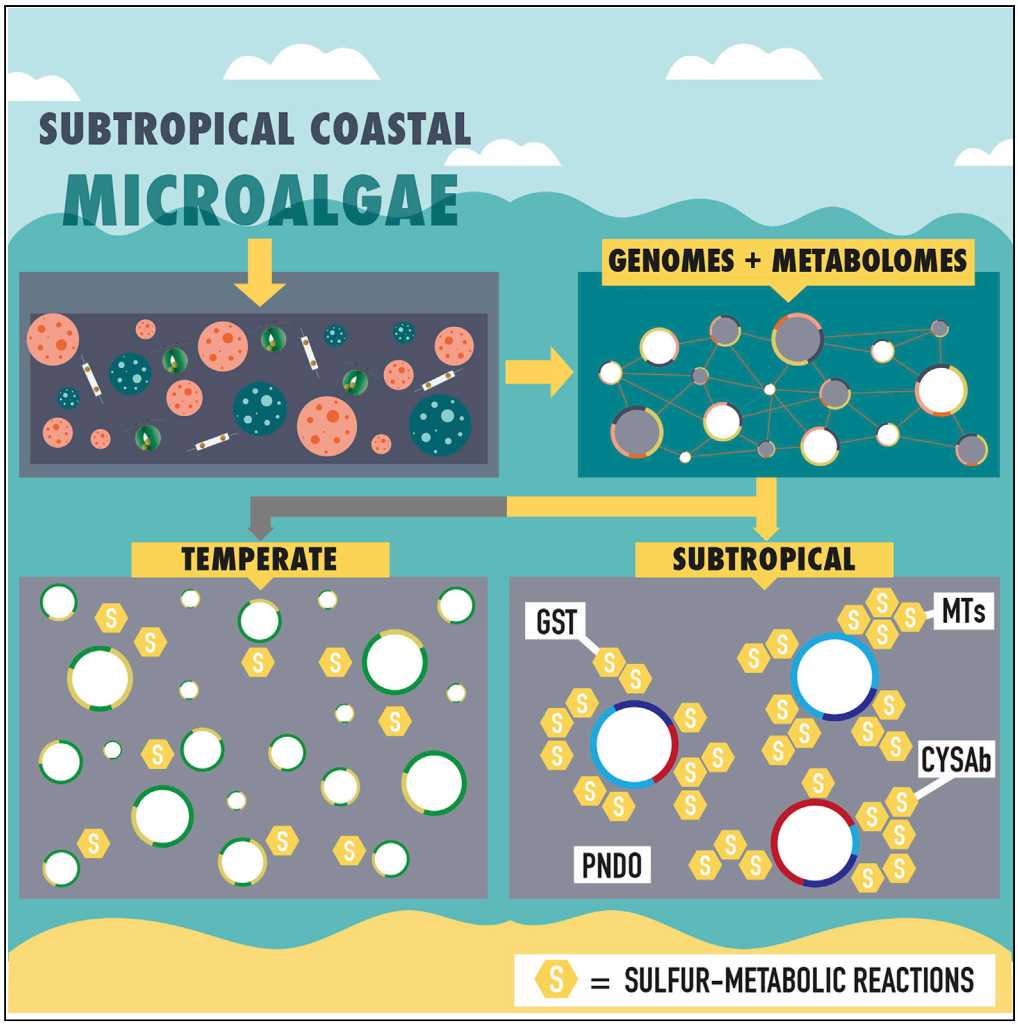
iScience

CellPress

Article

Potential for Heightened Sulfur-Metabolic Capacity in Coastal Subtropical Microalgae



David R. Nelson, Amphun Chaiboonchoe, Weiqi Fu, ..., Marc Arnoux, Mehar Sultana, Kourosh Salehi-Ashtiani

ksa3@nyu.edu

HIGHLIGHTS

We have sequenced 20+ microallgal genomes from the subtropics

This new collection increases the available microalgal genomes by ~50%

Metabolomics indicates lineage- and habitat-specificity of biomolecules

Coastal, subtropical species of microalgae show expansion of sulfur-metabolic genes

Nelson et al., iScience *11*, 450–465
January 25, 2019
https://doi.org/10.1016/j.isci.2018.12.035

iScience

CellPress

Article

Potential for Heightened Sulfur-Metabolic Capacity in Coastal Subtropical Microalgae

David R. Nelson,[1] Amphun Chaiboonchoe,[2] Weiqi Fu,[2] Khaled M. Hazzouri,[3] Ziyuan Huang,[4] Ashish Jaiswal,[2] Sarah Daakour,[1] Alexandra Mystikou,[1] Marc Arnoux,[1] Mehar Sultana,[1] and Kourosh Salehi-Ashtiani[1,2,5,*]

SUMMARY

The activities of microalgae support nutrient cycling that helps to sustain aquatic and terrestrial ecosystems. Most microalgal species, especially those from the subtropics, are genomically uncharacterized. Here we report the isolation and genomic characterization of 22 microalgal species from subtropical coastal regions belonging to multiple clades and three from temperate areas. Halotolerant strains including *Halamphora*, *Dunaliella*, *Nannochloris*, and *Chloroidium* comprised the majority of these isolates. The subtropical-based microalgae contained arrays of methyltransferase, pyridine nucleotide-disulfide oxidoreductase, abhydrolase, cystathionine synthase, and small-molecule transporter domains present at high relative abundance. We found that genes for sulfate transport, sulfotransferase, and glutathione S-transferase activities were especially abundant in subtropical, coastal microalgal species and halophytic species in general. Our metabolomics analyses indicate lineage- and habitat-specific sets of biomolecules implicated in niche-specific biological processes. This work effectively expands the collection of available microalgal genomes by ~50%, and the generated resources provide perspectives for studying halophyte adaptive traits.

INTRODUCTION

The environment of the United Arab Emirates (UAE) includes coastal deserts, mangroves, oases, and rocky valleys serving as basins for sparse precipitation (<75 mm/year). Coastal habitats, including mangrove forests, exhibit strong carbon sequestration capacity and contribute to the reversal of climate change effects (Ezcurra et al., 2016; Hader et al., 2014). Microalgae, both marine and terrestrial, inhabiting these areas are especially important participants in the carbon sequestration process (Ezcurra et al., 2016). Until recently, the Arabian Gulf received relatively low influx of essential nutrients, including nitrogen, phosphorous, and potassium (NPK). The rapid increase of human population in recent years and population-linked anthropogenic nutrient release have the potential to greatly affect the populations of microbial primary producers, such as microalgae (Paerl et al., 2016; Yang et al., 2018). Microalgae also play essential roles in the cycling of elements other than NPK, especially sulfur, with far-reaching, but understudied, consequences (Alcolombri et al., 2015; Giordano et al., 2005; Hell, 2008; Hurtgen, 2012; Pjevac et al., 2014; Wasmund et al., 2017). In this study, we focus on the role of sulfur metabolism in microalgae since the abundance of extracellular sulfur, in the form of sulfate, is one of the major factors that define the environment of marine microalgae, but not freshwater or terrestrial microalgae (Hurtgen, 2012). Sulfur can be limiting for most terrestrial species, whereas sulfur is abundant in marine environments because of the weathering of continental rocks (Hell, 2008; Hurtgen, 2012). However, the genomic differences between marine and freshwater microalgal species have not been considered, with respect to salinity and sulfur, previously because of the lack of sequenced genomes.

Historically, intensive research into the genic aspect of sulfur metabolism in microalgae has been limited to dimethyl sulfide (DMS) and its precursor dimethyl sulfopropionate (DMSP) (Haworth et al., 2017; Kettles et al., 2014; Raina et al., 2013; Sunda et al., 2002). Because DMS, a molecule known for its environmental cooling and precipitation-forming effects (Schafer et al., 2010), is a significant product from DMSP, the capacity of microalgae to engage in DMSP-related metabolism is especially relevant for hot and arid areas like the UAE. Two recent studies described previously unknown, essential, biosynthetic enzymes for DMSP. One study focused on a bacterial homolog that was present in several microalgal lineages, but not diatoms (Curson et al., 2018), and the other study focused on a eukaryotic homolog present in the diatom *Thalassiosira pseudonana* (Kageyama et al., 2018). By understanding the genomic basis for a single sulfur-based metabolic pathway, we are able to obtain new insight into the roles of microalgae in the global sulfur cycle. Thus, in this project, we aimed to illuminate potential sulfur-metabolic pathways in a broad

[1]Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, UAE

[2]Division of Science and Math, New York University Abu Dhabi, Abu Dhabi, UAE

[3]Khalifa Center for Genetic Engineering and Biotechnology (KCGEB), United Arab Emirates University, Al-Ain, UAE

[4]Department of Computer Science, New York University Shanghai, Shanghai, China

[5]Lead Contact

*Correspondence: ksa3@nyu.edu

https://doi.org/10.1016/j.isci.2018.12.035

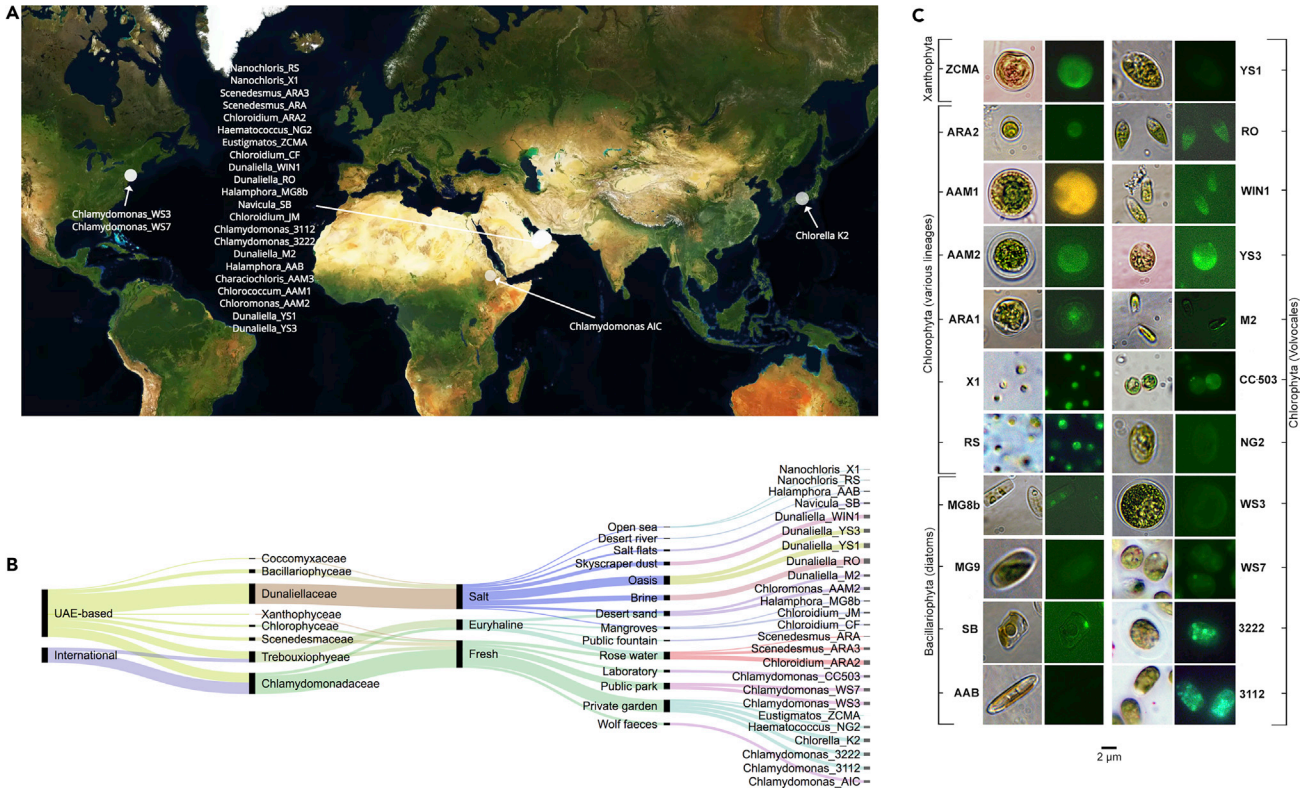


Figure 1. Isolation and Classification of Newly Isolated Microalgal Species

(A) Global locations of isolated microalgal species sequenced in this study.

(B) Alluvial flow diagram describing the newly isolated microalgal strains regarding their respective clades and isolation locations and environments.

(C) Morphology of isolates revealed by bright-field (left) and fluorescent (right) microscopy. Green autofluorescence, observed in right panels, marks the accumulation of intracellular compounds including, but not limited to, flavins, pigments, gaseous molecules, and photoactive proteins.

range of microalgae. Except for a few model species, microalgae in general, especially from subtropical, coastal areas, are not well characterized at the genomic level. This lack of knowledge surrounding their genic contents severely limits the ability to monitor their populations and ecosystem-level genetic interactions. In this study, we present the discovery, isolation, culture, whole-genome sequencing, assembly, annotation, and metabolomics of multiple new species of microalgae and, through comparative analyses, describe their unique biology and its potential impact on phycology, genomics, and biogeochemistry.

RESULTS

We isolated microalgal species from habitats including desert sands, mangroves, urban structures and gardens, open sea, public fountains, and desert oases to represent a range of natural and human-made environments (Table S1). The majority of isolates were chlorophytes from the UAE (Figure 1). Other isolates included several diatom species from salt flats, brine efflux outlets, mangroves, coastal areas, and shallow reefs. Many species were isolated from habitats that were at temperatures of 40°C or higher at the time of isolation (e.g., black salt muds at elevated temperatures), implying adaptation to high temperature and salinity. In the laboratory, we could cultivate most subtropical strains in agar-based media at $\geq$35°C; in contrast, their temperate counterparts, exemplified by *Phaeodactylum tricornutum* and *Chlamydomonas reinhardtii* for diatoms and chlorophytes, respectively, failed to grow at such temperatures. To create a resource for understanding microphytic life in the subtropics, we sequenced 22 new isolates from a variety of subtropical environments (Table S1).

Upon *de novo* assembly of the genomic reads obtained from the isolated species, we identified and aligned ribulose-1,5-bisphosphate-carboxylase/oxygenase (RbcL) genes from each new species and performed an alignment with RbcL genes from microalgae with publicly available genomes (Figure 2 Data

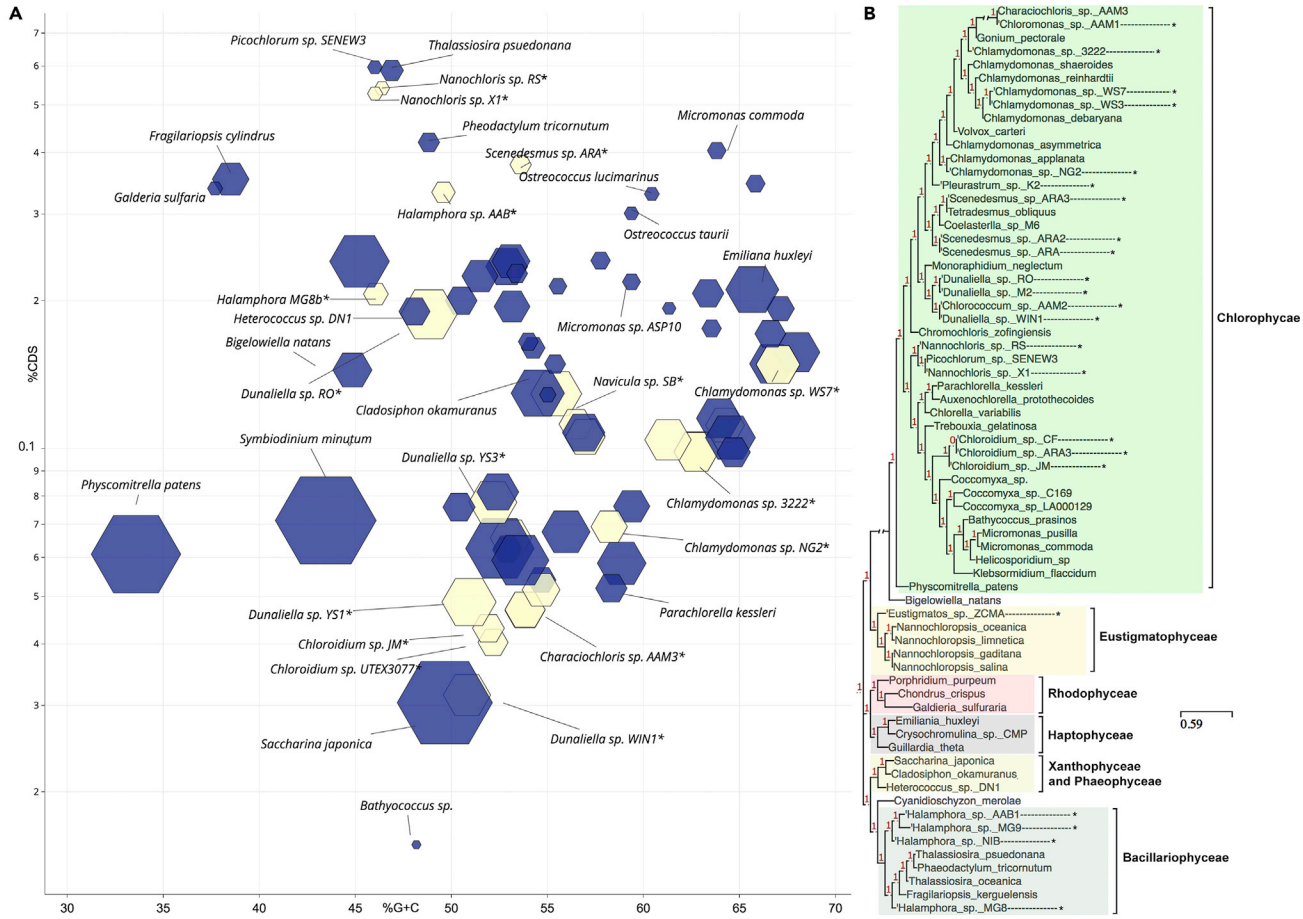


Figure 2. Genomes and RbcL-Based Phylogenetic Tree Reconstruction of the Isolates

(A) Range in size, %G + C content, and %CDS content among eukaryotic microalgae with sequenced genomes. Genomic assemblies presented in this study are marked with asterisks and described in more detail in Table S1. Genomic assemblies from our study (yellow) and NCBI (purple) are compared based on size (hexagon size), %G + C content (x axis), and %CDS (y axis, logarithmic scale). Genomic assembly metrics were obtained using QUAST (Gurevich et al., 2013).

(B) Alignment of ribulose-1,5-bisphosphate carboxylase/oxygenase (RbcL) genes and evolutionary model of species presented in this study and the genomic assemblies of microalgae available from NCBI. Asterisks mark the species that we sequenced. A value of "1" at the branch points indicates that the displayed evolutionary model had $p < 0.05$ when compared with the null model. We downloaded RbcL genes as annotated from their respective assemblies (Data S3). In the case that they were not available, we performed a BLASTn search using the *Chlamydomonas reinhardtii* RbcL gene (NCBI accession: NC_005353.1) against our predicted CDS set for the organism and extracted the top hit. Sequences, alignment, and tree are in Data S1, and an expanded tree that uses more species' RbcL genes are in Table S1.

S1). The Volvocales, including *Chlamydomonas* and *Dunaliella* species, were found to have the largest genomes of the green algae (Chlorophyceae), ranging from 80 to 140 megabase pairs (Mbp) with the highest %G + C contents among microalgae (Figure 2, Table S2). All the newly isolated and sequenced Volvocales species presented here were located in terrestrial environments within 20 km of the coastline. In particular, Volvocales species were readily isolated from most shaded terrestrial sites, suggesting that these may be the dominant, culturable microalgae from these areas. Coding sequences, including potential isoforms, from the newly isolated Volvocales species consisted of 30–60 Mbp and occupied a lower percentage of their respective genomes compared with marine microalgal strains (Figure 2, Table S2). In contrast, marine microalgae, including *Picochlorum*, *Nannochloris*, *Micromonas*, *Phaeodactylum*, and *Ostreococcus* species, had more compact genomes concerning their percentages of total transcriptome length (bp) to total genome length (bp) ratios (Table S2).

The marine strains could grow in media containing >40 g/L of NaCl. Although most strains preferred either salt or fresh water, the *Chloroidium* strains (*Chloroidium* sp. *CF* and *Chloroidium* sp. *JM*) that we sequenced

were euryhaline strains that thrived in both salt and fresh water. We isolated two species of the picoeukaryotic, free-living marine *Nannochloris* from the open waters of the Arabian Gulf. These two isolates represent culturable microalgae from sampling sites that were the furthest away from the coast into open waters and were the only picoeukaryotes sequenced in this study (Table S1). Picophytoplankton, consisting of microalgae with diameters ranging from 0.2 to 2 μm, are key primary producers in open ocean regions (Tragin and Vaulot, 2018). Populations of picoeukaryotic microalgae, such as *Nannochloris*, have been predicted to be among the winners in a future, acidified ocean because of their favorable responses to elevated CO_2 and decreased pH (Bach et al., 2017). Their small genome sizes (13–14 Mbp) and high protein-coding potentials (9–10 Mbp of predicted transcripts) display compact genomes with extensive intronation in coding regions (Figure 2).

Green microalgae, or chlorophytes, are also abundant in coastal waters (Tragin and Vaulot, 2018). The isolation of *Chloroidium* and *Dunaliella* species was especially prevalent in coastal sites, suggesting that these two Chlorophyta genera dominate the eastern coasts of the Arabian Gulf. Studies conducted as part of the World Ocean Sampling Day using operational taxonomic units from metagenomic data samples have estimated that as much as 42%–85% of occupant microalgae are chlorophytes in some European coastal areas (Tragin and Vaulot, 2018). Metagenomic datasets obtained from other sampling projects, such as the Tara ocean expedition (2009–2012), suggest that ochrophytes, including diatoms, are much more abundant in open oceans (de Vargas et al., 2015). Although the genera presented in this study have representations in many locations around the world, they were not among the most abundantly described in samples from the Tara or World Ocean Sampling Day (WOSD) (de Vargas et al., 2015; Tragin and Vaulot, 2018).

We collected records of known distribution for the reported taxa, and they show a wide distribution in different areas in Southwest Asia, which is the closest geographic region to the UAE, where the herein reported species were collected from (Table S3). The diatoms we isolated and sequenced were closely related to either *Navicula* or *Halamphora* species and had genomes ranging from 23 Mbp to 86 Mbp (Table S2). The isolation of *Halamphora* species from multiple inland locations, including mangroves, salt flats, rocky basins, and oases (Table S1), indicates that this clade has a broad domain throughout the region. Overall, the diatom coding sequences' transcripts had fewer introns per gene, whereas green microalgal species had more introns per gene (Data S2). Furthermore, species from the *Nannochloris* and *Chloroidium* genera had more introns, on average (four per gene), than species from the Volvocales clades (two per gene).

HMM-Based Functional Predictions and Biclustering of Species

We compared coding sequences (CDSs) from the UAE strains with those from all the publicly available microalgal genomes that we could access at the time of analysis (Figure 3). We also included *Physcomitrella patens*, *Saccharina japonica*, *Cladosiphon okamuranus*, and *Ectocarpus siliculosus* genomes in addition to the collection of 45 microalgal genomes (Data S3). We annotated all the genomes *de novo* to avoid variability in annotation accuracy from influencing the results (Data S4). The lower plant genomes act as an outgroup and help to provide contrast and perspective in our analyses regarding protein family and gene duplication events.

Our genomic annotations, based on hidden Markov model (HMM) alignments with amino acid sequences from experimentally verified proteins, were used for the reconstruction of gene structures and the detection of protein family domains (Pfams) in translated coding sequences. HMM matches identified a broad range of functional domains that the microalgal species contained differentially. Using sum confidence scores as indicators for the presence and copy number of functional domains, we carried out biclustering of the detected Pfam domains and their respective species (Figure 3). The outgroup included several multicellular lower plants and formed its primary cluster that also included phylogenetically distant microalgae from Arachinophyceae, Dinophyceae, Cryptophyceae, Xanthophyceae, and Rhodophyceae (Data S3). In general, freshwater- and saltwater-dwelling microalgae clustered within their respective groups and separate from each other based on standardized HMM-defined gene content (Figures 3 and 4). Table S2 describes each microalgal species used for our comparative analyses and whether we classified the species as saltwater, freshwater, or euryhaline.

Chlamydomonas species, including *C. reinhardtii* (*v5.5 assembly*) and other *Chlamydomonas* strains or species we isolated and sequenced (*Chlamydomonas* sp. *NG2*, *Chlamydomonas* sp. *3112*, *Chlamydomonas* sp. *3222*, *Chlamydomonas* sp. *WS3*, and *Chlamydomonas sp. WS7*), and other freshwater green

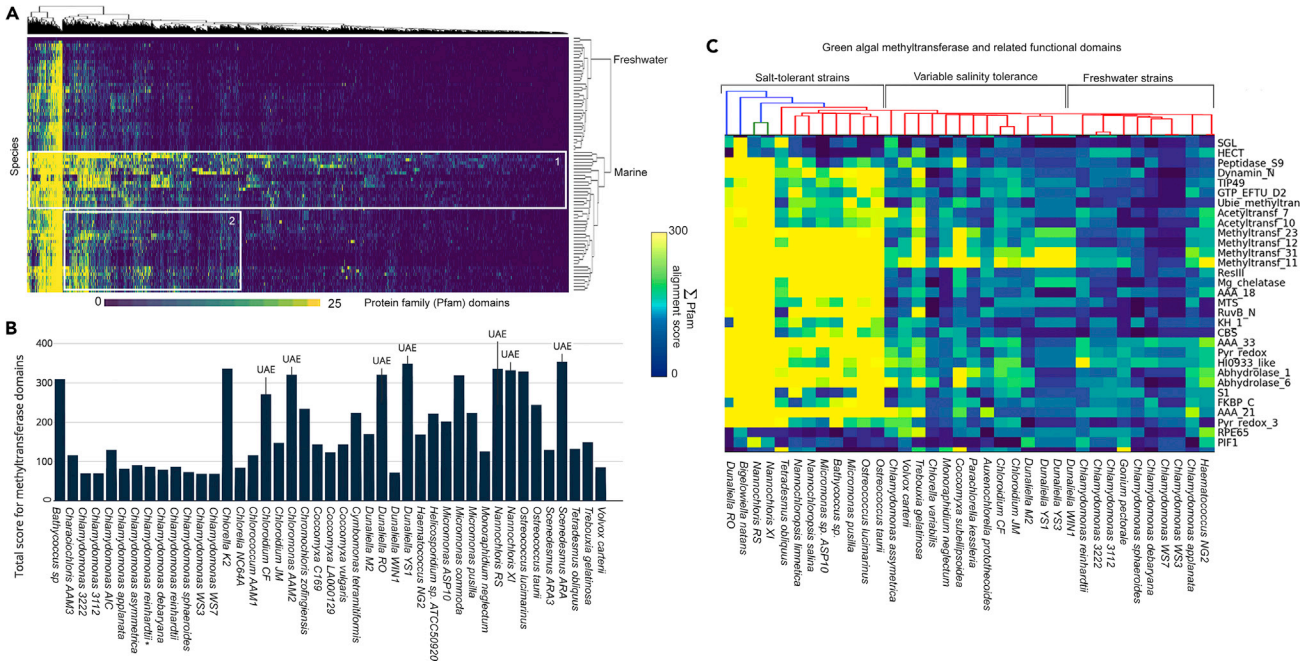


Figure 3. Protein Family Domain Analysis in New Microalgal Genomes and Microalgal Genomes Available in Public Databases

(A) Complete linkage hierarchical clustering based on the Euclidean distance of protein family domains (Pfams, x axis) and microalgal species (y axis) was performed in R using the heatmap.2 function in the gplot package. Z-depth (color) represents the sum score of hidden Markov model (HMM) alignments for their respective domains. Bounded box 1 represents Pfams from mostly outgroups, including multicellular or specialized, species. Box 2 represents Pfams enriched in marine species compared with freshwater species.

(B) An expansion of clustered domains from A, which were significantly enriched in marine and UAE-based species. These domains primarily consisted of methyltransferase domains.

(C) HMM alignment scores for all methyltransferase domains, and other domains, including abhydrolase domains, differentially abundant among Chlorophytes. UAE-based strains were among those with the highest copy number, and most significant matches for all methyltransferases, including the Methyltransf_11 domain that is known to interact with S-adenosyl-methionine.

microalgae *Chloroidium sp. UTEX3077* (Nelson et al., 2017), *Chlorella variabilis* NC64A, and *Coccomyxa subellipsoidea* clustered together, indicating a relatively similar HMM functional match profile (Figure 3). *Dunaliella* species were the exception to this clustering based on phylogenetic affinities, as they clustered with marine strains including diatom (Bacillariophyceae) and picoeukaryotic, oceanic (*Ostreococcus* and *Micromonas*) species.

Although microalgal species clustered tightly according to their clades and habitats in our Pfam analyses, biclustering based on EC (Enzyme Commission) assignments (Alborzi et al., 2017) within Kyoto Encyclopedia for Genes and Genomes (KEGG) pathways (Data S5) (Porollo, 2014) was inconsistent with Pfam-only biclustering, and species from the same clade did not always cluster together (Figure 4). This discrepancy between our Pfam analyses and KEGG pathway analyses could partially result from the distance each descriptor has to experimental evidence of function. For example, Pfam descriptors are directly based on experimental data and are frequently updated (Finn et al., 2016), whereas EC numbers are often indirectly related to empirical evidence and infrequently updated.

Our clustering results show that UAE-based strains, regardless of whether they were saltwater- or freshwater-preferring strains, share a significant number of functional domains with salt-tolerant and marine microalgal species. Because the UAE, on average, receives less than 75 mm of rainfall annually and has high soil salinity, these common gene suites may play essential roles in drought and salinity tolerance. We extracted some domains, including the methyltransferase (MT) domains shown in Figure 3C, and we detected a relative abundance in UAE-based species.

We found several groups of protein domains that were over-represented in the halotolerant species versus the freshwater species, and vice versa ($p < 0.00001$, Figures 3 and 5). These included, primarily, domains

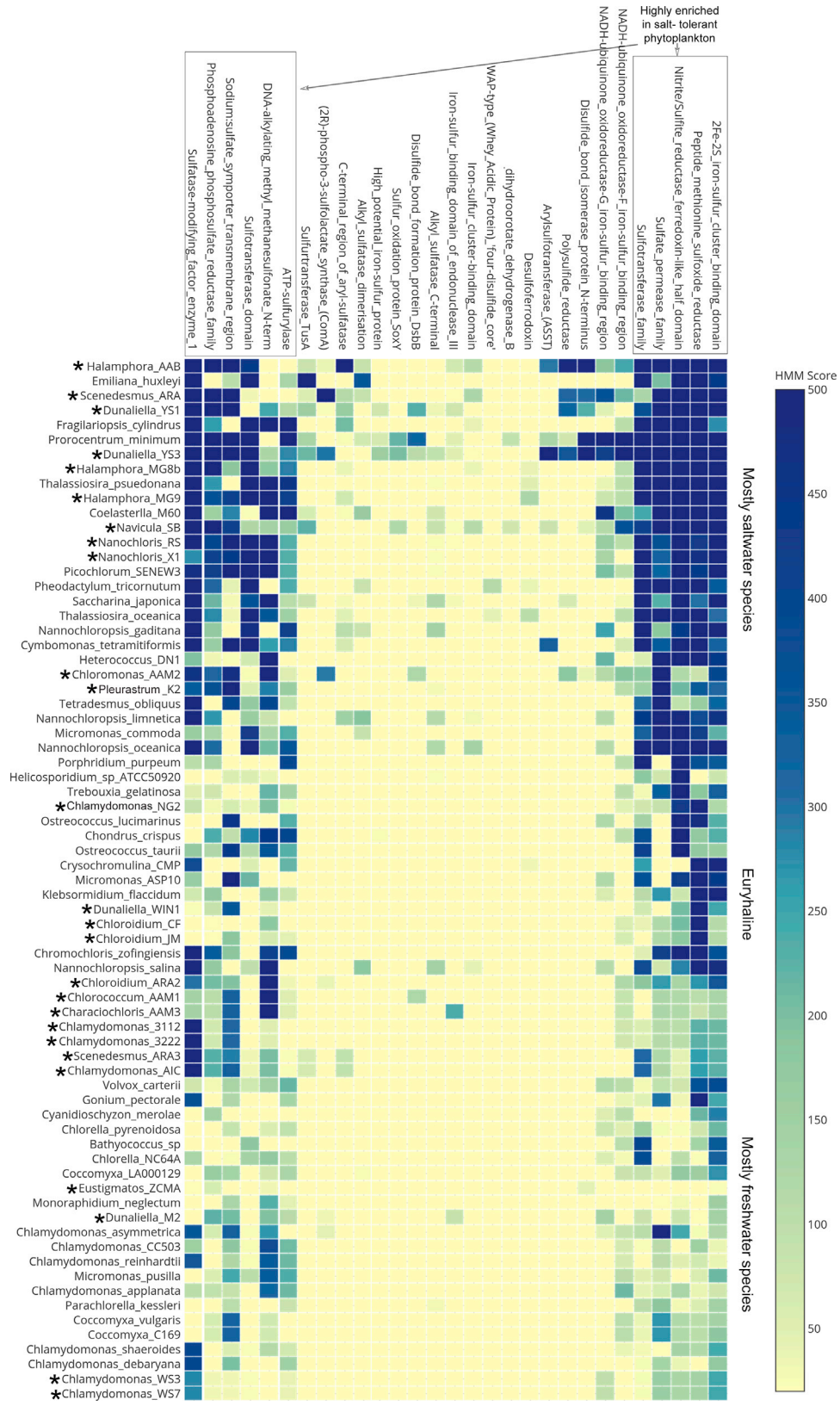

Highly enriched in salt- tolerant phytoplankton
2Fe-2S_iron-sulfur_cluster_binding_domain
Peptide_methionine_sulfoxide_reductase
Nitrite/Sulfite_reductase_ferredoxin-like_half_domain
Sulfate_permease_family
Sulfotransferase_family
NADH-ubiquinone_oxidoreductase-F_iron-sulfur_binding_region
NADH-ubiquinone_oxidoreductase-G_iron-sulfur_binding_region
Disulfide_bond_isomerase_protein_N-terminus
Polysulfide_reductase
Arylsulfotransferase_(ASST)
Desulfoferrodoxin
dihydroorotate_dehydrogenase_B
WAP-type_(Whey_Acidic_Protein)_'four-disulfide_core'
Iron-sulfur_cluster-binding_domain
Iron-sulfur_binding_domain_of_endonuclease_III
Alkyl_sulfatase_C-terminal
Disulfide_bond_formation_protein_DsbB
Sulfur_oxidation_protein_SoxY
High_potential_iron-sulfur_protein
Alkyl_sulfatase_dimerisation
C-terminal_region_of_aryl-sulfatase
(2R)-phospho-3-sulfolactate_synthase_(ComA)
Sulfurtransferase_TusA
ATP-sulfurylase
DNA-alkylating_methyl_methanesulfonate_N-term
Sulfotransferase_domain
Sodium:sulfate_symporter_transmembrane_region
Phosphoadenosine_phosphosulfate_reductase_family
Sulfatase-modifying_factor_enzyme_1
Halamphora_AAB
Emiliana_huxleyi
Scenedesmus_ARA
Dunaliella_YS1
Fragilariopsis_cylindrus
Prorocentrum_minimum
Dunaliella_YS3
Halamphora_MG8b
Thalassiosira_psuedonana
Halamphora_MG9
Coelasterlla_M60
Navicula_SB
Nanochloris_RS
Nanochloris_X1
Picochlorum_SENEW3
Pheodactylum_tricornutum
Saccharina_japonica
Thalassiosira_oceanica
Nannochloropsis_gaditana
Cymbomonas_tetramitiformis
Heterococcus_DN1
Chloromonas_AAM2
Pleurastrum_K2
Tetradesmus_obliquus
Nannochloropsis_limnetica
Micromonas_commoda
Nannochloropsis_oceanica
Porphridium_purpeum
Helicosporidium_sp_ATCC50920
Trebouxia_gelatinosa
Chlamydomonas_NG2
Ostreococcus_lucimarinus
Chondrus_crispus
Ostreococcus_taurii
Crysochromulina_CMP
Micromonas_ASP10
Klebsormidium_flaccidum
Dunaliella_WIN1
Chloroidium_CF
Chloroidium_JM
Chromochloris_zofingiensis
Nannochloropsis_salina
Chloroidium_ARA2
Chlorococcum_AAM1
Characiochloris_AAM3
Chlamydomonas_3112
Chlamydomonas_3222
Scenedesmus_ARA3
Chlamydomonas_AIC
Volvox_carterii
Gonium_pectorale
Cyanidioschyzon_merolae
Chlorella_pyrenoidosa
Bathyococcus_sp
Chlorella_NC64A
Coccomyxa_LA000129
Eustigmatos_ZCMA
Monoraphidium_neglectum
Dunaliella_M2
Chlamydomonas_asymmetrica
Chlamydomonas_CC503
Chlamydomonas_reinhardtii
Micromonas_pusilla
Chlamydomonas_applanata
Parachlorella_kessleri
Coccomyxa_vulgaris
Coccomyxa_C169
Chlamydomonas_shaeroides
Chlamydomonas_debaryana
Chlamydomonas_WS3
Chlamydomonas_WS7
Mostly saltwater species
Euryhaline
Mostly freshwater species
HMM Score
500
450
400
350
300
250
200
150
100
50

Figure 4. Elevated Sulfur-Related Protein Families (Pfams) in Salt-Tolerant Algae

Sulfur-metabolic domains in salt-tolerant algae showed significantly higher sum confidence scores than phytoplankton from the other areas. Asterisks mark species that we present in this article as new genomic assemblies.

involved in osmoregulation, oxidative stress resistance, sulfur metabolism, and carbon metabolism. We also found some protein families over-represented in the subtropical-based strains, compared with strains from other, more temperate areas. Specifically, the UAE-based microalgae, including *Nannochloris sp. XI*, *Nannochloris sp. RS*, *Dunaliella sp. YS1*, *and Dunaliella sp. YS3*, contained arrays of MT, pyridine nucleotide-disulfide oxidoreductase, abhydrolase, cystathionine synthase, and small-molecule transporter domains present in higher copy numbers or unique to both freshwater and saltwater subtropical species and other marine species (Figure 3, Table S3).

Sulfur-Metabolic Gene Abundance in Marine and Subtropical Microalgae

Genes responsible for sulfur-based metabolisms, including sulfate transport, and cysteine- and methionine-related genes were expanded in subtropical and marine microalgae, as observed from their HMM match profiles (Data S6) (Figures 5 and 6). The expansion of sulfate transport capacity might allow for the influx of additional sulfur needed for intracellular molecules including methionine, cysteine, glutathione, S-adenosyl methionine, biotin, coenzyme A, thiamine, etc. Recently, two groups described the fast-acting antioxidant role of N-acetylcysteine, which is directly limited by sulfur supply to the organism (Cerda and Pluth, 2018; Ezerina et al., 2018). Thiol-related domains, including glutathione S-transferase (GST), were relatively abundant in salt-tolerant strains, although not limited to UAE-based strains (Figure 5).

We observed an abundance of MT domains in subtropical coastal and marine microalgae (Figure 3C). Several MT domains were included in relative abundance in marine species (Figure 3B); however, the Methyltransf_11 domain was found in high copy number in both marine species and terrestrial halophytes, including *Dunaliella* species. Overall, our data suggest that the abundance of sulfur-metabolism genes (Figure 4) is a primary feature of the subtropical species we isolated, cultured, and sequenced. We saw it particularly pronounced in the species that dominated our environmental isolations, namely, *Nannochloris*, *Halamphora*, and *Dunaliella* species.

The expanded sulfur metabolism in these species and other salt-tolerant microalgae may be due to the increased availability of inorganic sulfur in marine environments in recent and prior geological eras (Giordano et al., 2005) and the necessity to cope with salt stress. The expanded GST sequences we observed in marine species may be a result of utilization of sulfur as well as the high salinity in marine environments, and the restriction of GST sequences in terrestrial species could be due to limited sulfur availability as well as the absence of salt stress. We also observed increased abundance of related encoded protein domains that may function interdependently in metabolic networks. For example, in the case of methionine synthase and GS, methionine biosynthesis is limited by the availability of sulfur; however, cobalamin, which is an obligate cofactor for one isoform of methionine synthase, has been shown to require GST activity to process alkylcobalamins into their active cofactor forms (Kim et al., 2009).

Oxidative stress tolerance has been shown to be tightly linked to salt tolerance in higher plants (Siddiqui et al., 2017), and several of the most powerful cellular antioxidant systems are sulfur based, using a range of molecules (Battin and Brumaghim, 2009; Mukwevho et al., 2014). A higher amount of dissolved inorganic sulfur in oceanic waters could be incorporated into the cysteine residues of redox-active molecules, including glutathione, thioredoxin, and glutaredoxin, thus promoting the development of sulfur-based antioxidant systems able to withstand salinity and other abiotic stressors, such as constant solar radiation, that characterize marine ecosystems. Strong support for the role of sulfur biomolecules in salt stress mitigation is provided by the finding that the lipid peroxidation product, 4-hydroxynonenal, functionally interacts with GST (Balogh and Atkins, 2011). Thus, the increased presence of GST homologs in the saltwater strains may primarily serve to mitigate the damage caused to membrane lipids, DNA, and other intracellular molecules by salt stress.

Analysis of MTHB-MT Gene Homologs across Sequenced Diatoms

As the biosyntheses of DMSP and its product, DMS, have recently received much attention because of their implication in climate change processes (Kettles et al., 2014), we investigated the presence of enzymes involved in these processes in the newly isolated and sequenced microalgal species. Using BLASTn search

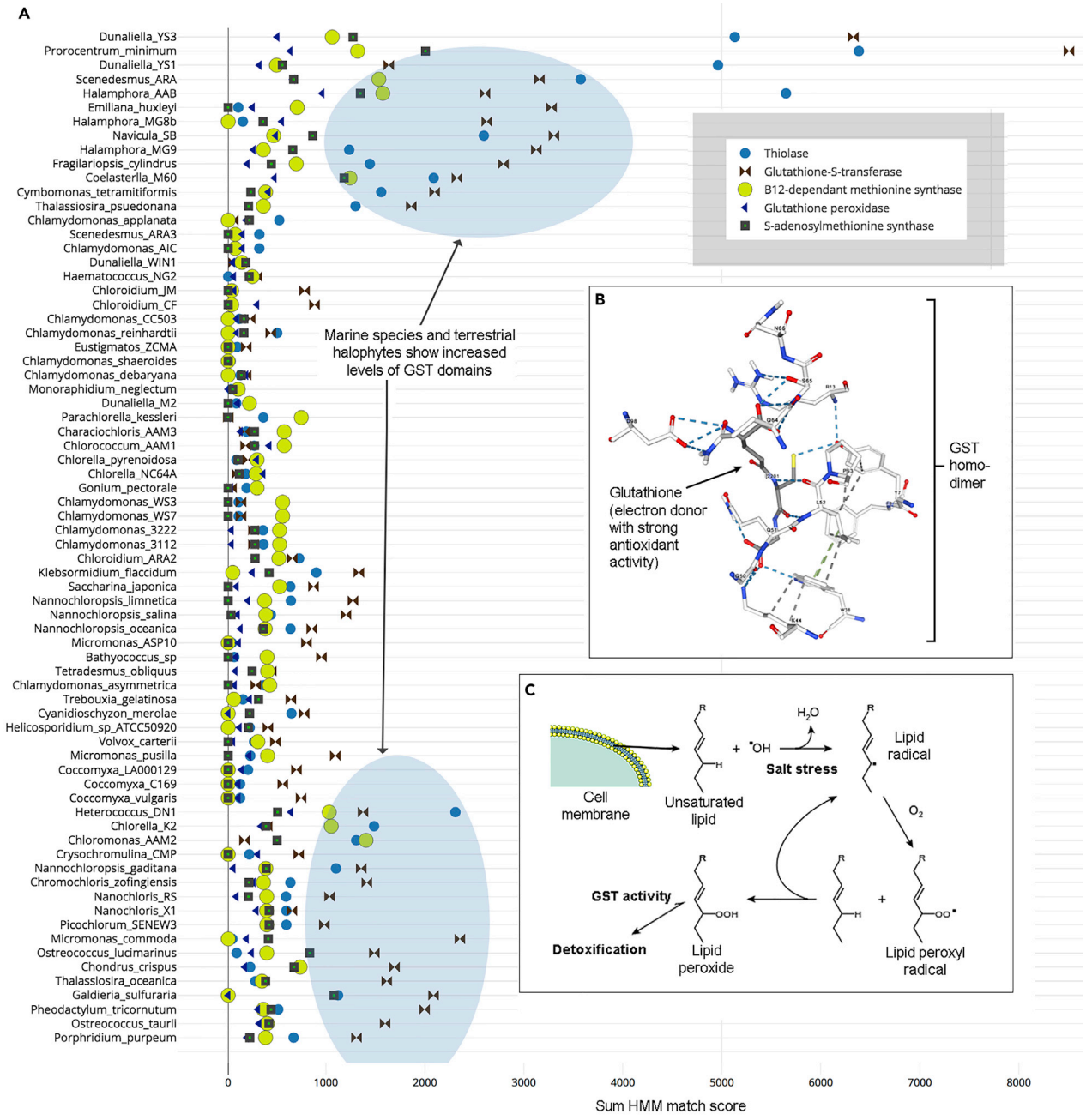


Figure 5. Thiol-Related Pfams in Green Microalgal Genomes and Their Possible Role in Salt Stress Tolerance

(A) The x axis represents sum Pfam confidence scores, where a score of greater than 100 indicates a strong probability that at least one true functional domain is present. Marine species, including *Ostreococcus* and *Micromonas*, had an unusually high copy number of glutathione-S-transferase (GST) domains. Terrestrial halophytes, including *Halamphora* and *Dunaliella* species, also shared this elevation in GST domain presence.

(B) Protein DataBank (PDB) model for glutathione-S-transferase (http://www.rcsb.org; IAQW; EC: 2.5.1.18; DOI: 10.2210/pdb1AQW/pdb; Prade et al., 1997; Rose et al., 2018) binding to the antioxidant molecule glutathione (GSH) from a representative homodimer form. GST enzymes can also coalesce as heterodimers.

(C) Proposed mechanism of membrane lipid oxidation, peroxide formation and propagation, and GST-mediated detoxification, based on studies performed by Balogh and Atkins (2011) showing detoxification of lipid peroxides mediated by GST activity; pathway diagram adapted from a .png file from Young and McEneny (2001) and the cell membrane graphic was adapted from Servier Medical Art, and the use of both graphics are covered under Creative Commons Attribution 3.0.

through the CDSs and a BLASTp search using proteins (Data S4), we were unable to find any homologs of DMSP-lyase, the enzyme responsible for DMS formation, in our dataset from our sequenced microalgae and the downloaded genomes that passed our strict confidence filter (E-value < 1e-10). However, we did observe homologs for genes essential to DMSP biosynthesis, including a methylthiohydroxybutyrate methyltransferase (MTHB-MT).

We analyzed the presence of MTHB-MT (Figure 6A), which is essential for the biosynthesis of DMSP, in three diatoms from the UAE (*Navicula sp. SB*, *Halamphora sp. AAB1*, and *Halamphora sp. MG9*) and three diatom genome assemblies downloaded from NCBI (*P. tricornutum*, *Fragilariopsis cylindrus*, *T. pseudonana*; Figure 6, Data S7). We found multiple BLASTp hits per species using the *T. pseudonana* MTHB-MT protein as a query (Kageyama et al., 2018). Most BLASTp hits and the MTHB-MT query showed a strong HMM match for the Methyltransf_11 Pfam domain (Table S4). The HMM for this specific MT domain, Methyltransf_11, or PF08241, has been characterized an S-adenosyl methionine (SAM)-dependent MT. Normally, SAM-dependent MTs use SAM to methylate proteins, small molecules, lipids, and nucleic acids (Martin and McMillan, 2002). All SAM-dependent MTs contain a structurally conserved SAM-binding domain, and these sequences can be used to build a type-specific HMM. In this case, the conserved domain is a central, seven-stranded β sheet surrounded by three α helices (Martin and McMillan, 2002).

In addition to the Pfam and BLASTp analyses, we also performed a MUSCLE (Edgar, 2004) alignment of the amino acid sequences and InterPro searches for matches (https://www.ebi.ac.uk/interpro) to elucidate whether the recovered diatom amino acid sequences were likely to be functional MTs (Data S7). Our MUSCLE alignment showed conservation of many residues within the predicted MT domain regions (Figures S1 and S2, Data S7). Searches through the InterPro database (as of October 2018) yielded strong hits with these proteins for SAM-binding domains as well as MT activity (InterPro results, including scores and annotations, can be found as .xml, .gff, and .tsv files in Data S7). InterPro also predicted significant conservation with extant, functional MTs and proteins that had between 6 and 12 conserved residues within their SAM-binding sites (Data S7).

In particular, all proteins within the alignment had 100% conservation of six residues consisting of a pair of phenylalanine and aspartic acid residues flanked on both sides by two pairs of glycine residues (Figure S2). These conserved residues are important in forming known secondary structures for MTs in general and especially the shared Methyltransf_11 domain, indicated in Figure S1. For example, β sheets form around aromatic residues like phenylalanine, whereas glycine pairs mark the borders of the central β-sheet fold in the active site. SAM-dependent MTs are known for these glycine-rich regions (Cooke et al., 2009). The β sheet is normally flanked by α helices or disordered amino acid chains, and these helix/disordered flanking sequences appear to be variable and displayed much less homology among our set of candidates (Figure 7 and S1). Low overall homology is another common feature of MTs (Cooke et al., 2009). In addition, although the aromatic residues and flanking glycines are common in most MTs, the aspartic acid we observed to be universally conserved in our candidate MTHB-MTs may play a specific role in acting on MTHB. Other phenol-flanking aspartic acid residues in unrelated MTs have also been postulated to aid in substrate activation, proton shuttling, and substrate specificity (Cooke et al., 2009).

Based on our BLASTp, Pfam, InterPro, MUSCLE, and RAxML results, we propose that the Methyltransf_11 domain is likely the active site responsible for the MT step in the biosynthesis of DMSP. Although the *P. tricornutum* MTHB-MT homolog (Figure 6) had both Methyltransf_31 and Methyltransf_11 domains, we only observed significant conservation in the Methyltransf_11 domain. Other MTHB-MT homologs contained Methyltransf_31, Methyltransf_23, and Methyltransf_25 domains either in addition to or in place of the Methyltransf_11 domain (Data S7, Figure 7).

Metabolomics of 14 Representative Strains and Cross-Lineage Comparisons

We performed liquid chromatography coupled with quadropole time-of-flight (QToF) mass spectrometry (LC-MS)-QToF on extracts of 14 representative microalgal strains to corroborate our findings of diversity between clades, and within clades, from different habitats (Figure 7; Data S8). Our method was calibrated to quantify both hydrophobic and hydrophilic compounds with a mass-to-charge ratio (*m/z*) ranging from 125 to 2,100. We identified compounds based on their ion chromatogram (elution time, minutes) and fragmentation patterns from MS-QToF. We observed unique and shared metabolites among these strains, and the complete results, including total ion chromatograms (TICs) before retention time alignment, retention time deviations, TIC

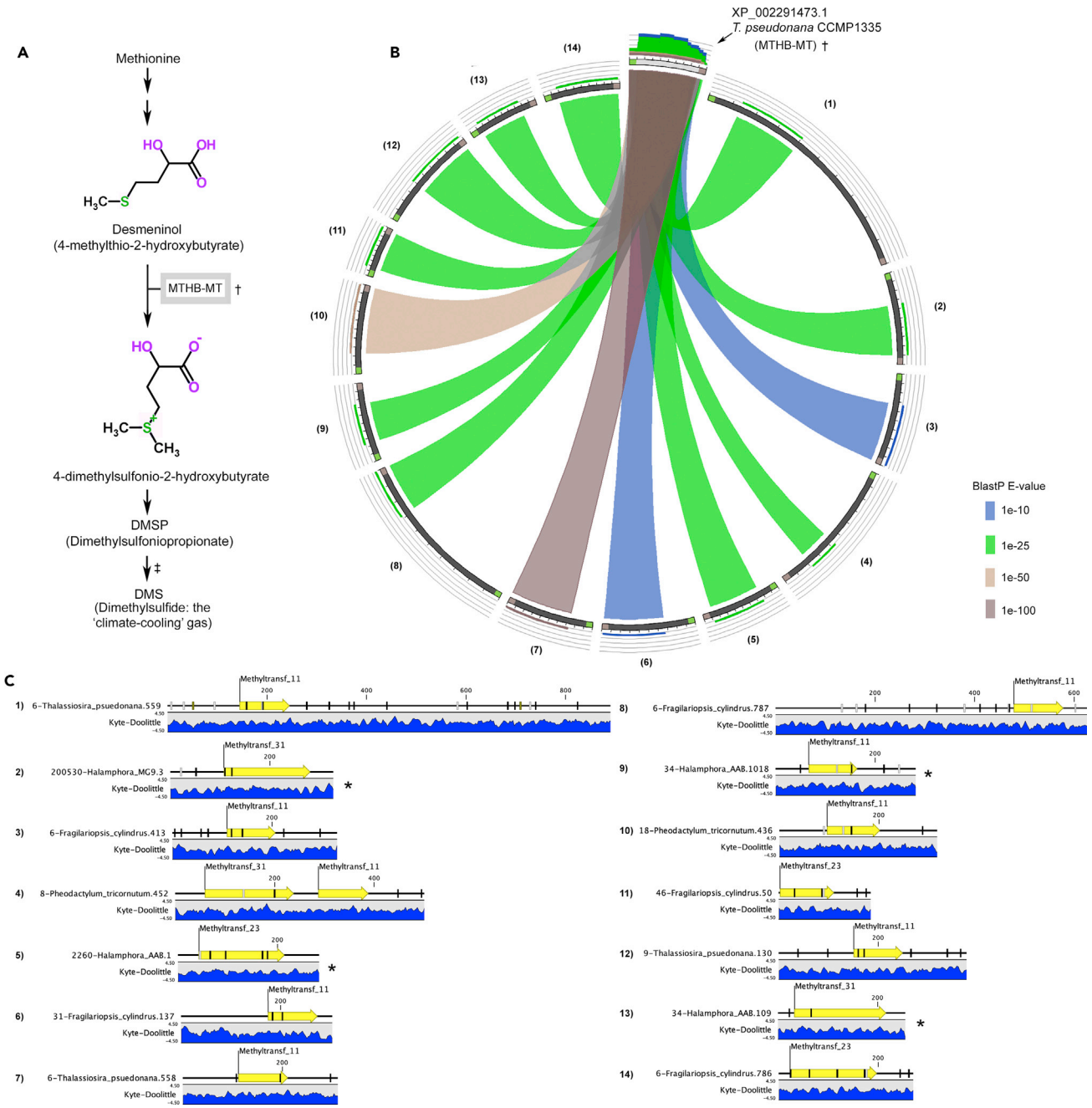


Figure 6. Circoletto (tools.bat.infspire.org/circoletto/) Plot of Alignment of CDSs within Diatoms

(A) Biosynthetic pathway for DMSP and DMS. The † symbol indicates the methylthiohydroxybutyrate methyltransferase (MTHB-MT) presented in Curson et al. (2018). The ‡ symbol indicates the DMSP lyase reaction known to occur in eukaryotic phytoplankton (Arnold et al., 2013) and bacteria (Lei et al., 2018).

(B) This alignment shows BLASTp hits (E < 1e-10) with a methylthiohydroxybutyrate methyltransferase (MTHB-MMT) that catalyzes an essential step in DMSP biosynthesis (Kageyama et al., 2018). The plot was made using Circoletto (tools.bat.infspire.org/circoletto/). Translated coding sequences were from our sequenced phytoplankton and publicly accessioned assemblies. Diatom genomes downloaded from NCBI included *Phaeodactylum tricornutum*, *Fragilariopsis cylindrus*, and *Thalassiosira pseudonana*. Diatom genomes sequenced by us and used for this study included *Halamphora sp. AAB*, *Halamphora sp. MG9*, and *Navicula sp. SB*. Although we observed one BLASTp hit within the *Navicula sp. SB* predicted proteins, the E-value for this hit was above the 1e-10 cutoff (1×10^{-6}), and we did not include it in the above diagram.

(C) Protein models corresponding to the aligned proteins. Methyltransferase domains, assumed to be the active enzymatic site for the methyltransferase activity, are annotated with yellow arrows. We indicate glycosylation sites as white bars and predicted phosphorylation sites as black bars along the amino acid sequences. Kyle-Doolittle plots, indicating hydrophobicity, are shown below each protein model. We indicate amino acid residues in the numbers on

Figure 6. *Continued*
the protein models. For a further expansion on TpMMT BLASTp hits from diatom proteins, see Data S7. A graphical MUSCLE alignment of these amino acid sequences highlighting conserved residues and gap fractions can be found in Figure S2, and InterPro search results in .xml format, as well as graphical .png files, can be found in Data S7.

post retention time correction, metabolite plots, multidimensional scaling, and principal-component analyses, are in Data S8. More than 19,889 metabolites were significantly different among strains ($p < 0.01$, Data S8).

To examine the diversity between lineages and between habitats, we categorized species as (1) diatoms, (2) saltwater chlorophytes, and (3) freshwater chlorophytes. For each molecular feature observed in our LC-MS-QToF study, we pooled the molecular counts (average of three biological replicates for each species) from three species for each group. For the freshwater chlorophytes, we used *Pleurastrum (Chlorella) sp. K2*, *Scenedesmus sp. ARA*, and *Chlamydomonas sp. A1E*. For the saltwater chlorophytes, we used *Dunaliella sp. RO*, *Dunaliella sp. YS3*, and *Dunaliella sp. YS1*. For the diatom group, we used *Halamphora sp. MG8*, *Halamphora sp. AAB*, and *Navicula sp. SB*. Although *Dunaliella* species are chlorophytes, they have a high degree of halotolerance and inhabit extremely different environments than their freshwater counterparts. The freshwater chlorophytes represent strains from terrestrial habitats including non-saline soils and bodies of fresh water. The diatoms and saltwater chlorophytes represent strains from marine and other saline environments, including salt flats. From this analysis, we were able to elucidate metabolites common to all strains and lineages and those metabolites that were specific to each of the three groups (Figures 8 and 9). Of these three groups, metabolites from the diatoms were more different from those of both the freshwater and saltwater chlorophytes than the chlorophytes from different salinities were to each other.

The ternary plot in Figure 7A allows us to observe the variance between two groups while controlling for the third group. In this instance, we can elucidate molecules that are common across lineages irrespective of the habitat and molecules that are common across habitats (salt water versus fresh water) irrespective of the phylogenetic affinity. We observed 339 compounds that were accumulated both in diatoms and saltwater chlorophytes, but not accumulated in freshwater chlorophytes. Of these 339 compounds, we observed several sulfur-containing molecules with biological roles that have not yet been well investigated.

The dataset of unique diatom metabolites (174 compounds) in Figure 7B represents a set of biomarkers for diatoms, including very-long-chain-fatty-acid-containing membrane lipids and an array of secondary metabolites. We observed that 3-methylthiopropionic acid (3MTPA) was specific to the diatom group, although this compound is a known intermediate in methionine metabolism in humans (Steele, 1978).

DISCUSSION

Coastal areas, including mangrove and other wetlands, comprised most of the areas of microalgae sampling in our study that can be extremely important in regional and global biogeochemistry (Ezcurra et al., 2016; Maclean and Wilson, 2011). However, nutrient deprivation and climate change threaten coastal environments in arid, subtropical regions like the UAE (Almahasheer et al., 2017). These sensitive areas have the potential to be highly active carbon sinks that could help offset atmospheric carbon discharge (Bose and Satyanarayana, 2017; Ezcurra et al., 2016; Seneviratne, 2003); however, before this study, little was known about their microbial primary producer constituents, particularly their genomic makeup in contrast to their temperate zone counterparts.

One limitation to our study is that we cannot wholly resolve differences within clades due to latitude or environment, such as subtropical versus temperate, in specific clades because no genomic data are currently available for species from these clades. Specifically, genomes of *Dunaliella*, *Nannochloris*, and *Halamphora* species have not been previously sequenced. However, isolates from these clades were abundant in our isolations and can serve as representative microalgal species from coastal, subtropical areas. In particular, we saw stark differences in many sulfur-metabolic genes in microalgae from different habitats, including their expansions in many coastal, marine, and terrestrial halophyte lineages, as visualized in Figures 3, 4, and 5. We found that GST domains were abundantly present in marine and terrestrial halophytes (Figure 5). In cellular sulfur metabolism, GSTs quench reactive molecules with the conjugation of glutathione (GSH) and detoxify cells from oxidative damage (Hell, 2008; Khan et al., 2013; Kumar and Trivedi, 2018). GSTs are involved in the detoxification mechanism, and 15 GST-like isoenzymes (*Cr*GSTs) were found in the model species of green algae *C. reinhardtii* (Chatzikonstantinou et al., 2017). Transgenic GST, introduced

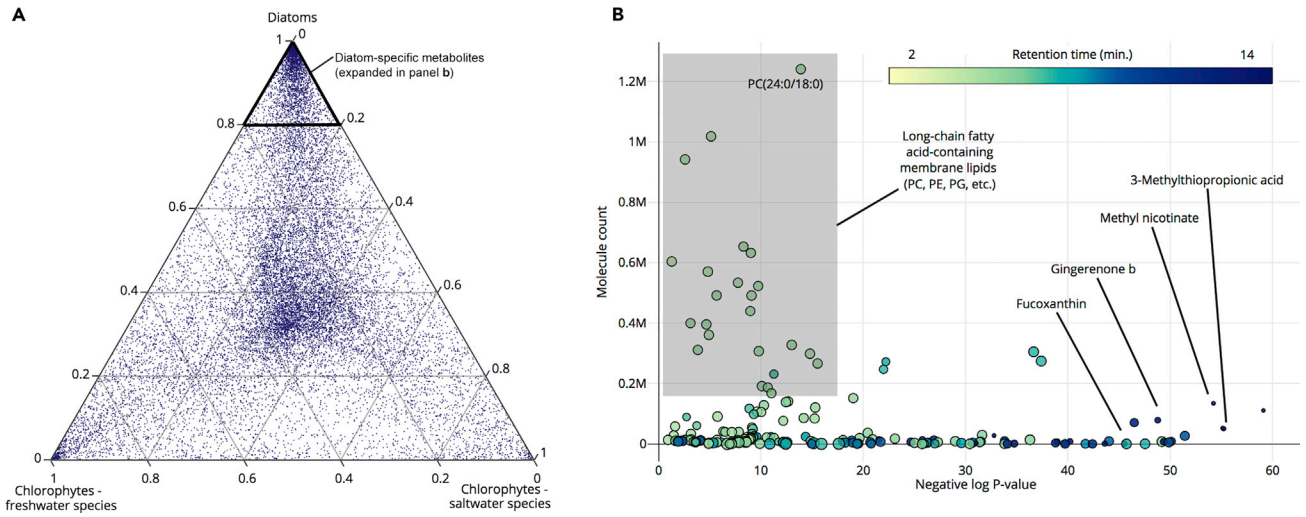


Figure 7. Extracted Metabolites from Representative Microalgae Species Highlight Diversity among Lineages (Diatoms versus Green Algae) and within Lineages Occupying Different Habitats (Freshwater Green Algae versus Saltwater Green Algae)

We performed metabolomics on 14 representative microalgal species. Species and their countries of isolation were *Dunaliella sp. YS3* (UAE), *Dunaliella sp. YS1* (UAE), *Chlamydomonas sp. AIE* (ETH), *Navicula sp. SB* (UAE), *Dunaliella sp. RO* (UAE), *Halamphora sp. MG8* (UAE), *Scenedesmus sp. ARA* (UAE), *Chlorococcum sp. AAM1* (UAE), *Eustigmatos sp. ZCMA* (UAE), *Halamphora sp. MG1* (UAE), *Chlorella sp. LF* (PT), *Pleurastrum sp. K2* (JP), *Chloromonas sp. AAM2* (UAE), and *Halamphora sp. AAB1* (UAE). All species were cultured in either freshwater or saltwater media for 1 month and underwent extraction using sonication followed by microwave-assisted methanol extraction and filtration.

(A) Ternary plot made from pooling the average count for each detected molecular feature (n = 20,618) from three strains for each group. We highlighted metabolites that were uniquely abundant in diatom species with the top, bold-outlined triangle.

(B) Molecules that were only contained within diatoms and not the other two groups in A (174 molecules). Statistically significant differences in metabolite abundance were found using two-tailed t tests, and we provide p values adjusted for a false discovery rate of 0.05 (Data S8). Fucoxanthin, a brown pigment, is known to accumulate exclusively in diatoms. However, 3-methylthiopropionic acid is a known methionine metabolism intermediate in other organisms, including humans (Steele, 1978), but among the microalgae in our metabolomics studies, we only observed diatoms accumulating this toxic molecule.

into *Arabidopsis thaliana*, has been shown to increase its salt tolerance (Xu et al., 2015). Owing to the increasing soil salinity and the desire to expand the salinity range for crop species, the new GST genes presented in this article are excellent candidates for transgene expression studies in higher plants. Because one shared element of these species' environments is the abundance of extracellular salts, and glutathione-related proteins have known roles in combating oxidative stress brought about by salt stress, these functionally conserved domains likely played substantial roles in the adaptation of their host to their respective saline environments. Transgenic GST has also been shown to promote tolerance to heavy metals, including cadmium (Dixit et al., 2011). The new array of GST homologs from the new isolates may also play roles in tolerance to heavy metals and are thus candidates for crop improvement or bioremediation studies.

The accumulation of sulfur-based antioxidant molecules, which may play even more important roles in heat tolerance when water is lacking, is not wholly characterized (Li et al., 2013). Although many genes that constitute similar functional metabolic networks are co-conserved (Chaiboonchoe et al., 2016), a few sulfur-based biomolecules are known to have overarching roles in environmental stress resistance and deserve a more intensive focus. Glutathione (Wilson et al., 2014) and DMSP (Kettles et al., 2014) are known to be crucial sulfur-based molecules for the resistance of oxidative stress caused by hot, arid conditions. The gene encoding MTHB-MT was recently reported to occur in marine algae (Curson et al., 2018), and its function is essential for DMSP biosynthesis. We, as in Kageyama et al. (2018), were able to readily detect genomic evidence for MTHB among diatom genomes, which is consistent with the reported production of DMSP in diatoms, even though a DMSP lyase is yet to be identified (Schafer et al., 2010) (Schafer et al., 2010).

Saline bodies of water disperse sulfur-containing molecules, among other minerals, necessary for both marine- and terrestrial-based life through degradation of sediment-containing rocks (Hell, 2008; Hurtgen, 2012). Thus halotolerant microalgae have an environmental basis for the retention and duplication of genes

like the sulfate transporters, sulfotransferase, and sulfatase enzymes we show in Figure 4, and also like the GST homologs we show in Figure 5. For example, glutathione and many other intracellular antioxidant mechanisms depend on and are limited by extracellular sulfur availability (Brose et al., 2014). We also note that diatoms use reduced sulfur compounds in the assimilation of extracellular silica (Lewin, 1954). We hypothesize, based on high-confidence functional HMM matches, that marine-based and other halotolerant microalgae have an expanded capacity to metabolize sulfur-containing compounds and that this expanded capacity is likely both necessitated and facilitated by their environment.

Most studies on DMS production, through DSMP lyase activity, have focused on bacterial sources (Schafer et al., 2010). DMSP lyases catalyze the breakdown of DMSP into DMS and acrylic acid (Lei et al., 2018). However, a recent study identified and characterized a DMSP lyase reaction in the eukaryotic species *Emiliana huxleyi* (Alcolombri et al., 2015). We did not observe orthologs to DMSP lyases in our studies at strict confidence filter levels, and this suggests that the breakdown of DMSP might be restricted to haptophytes and bacteria associated with DMSP-producing bacteria. Alternatively, the encoding gene may have diverged significantly; therefore, it could not receive a high confidence score in an HMM scan.

We found several unique sulfur compounds among the newly isolated microalgae. For example, we detected methyl-sulfanyl naphthalene at high levels in the *Halamphora* species. Naphthalenes have been previously shown to undergo oxidation in many photosynthetic organisms (Fassioli et al., 2014), and their photochemistry closely mimics that of chlorophyll molecules (Kopelman, 1976). The knowledge regarding the range of occurrences among different clades is sparse (Ibrahim and Mohamed, 2016); however, among the species that we examined, we observed naphthalene carboxamide in *Navicula sp. SB*, 2-nonylnaphthalene in *Halamphora* species, and methylnaphthalene molecules in varying degrees in every microalgal strain tested (Data S8). Considering their abilities to mimic known photosynthetic processes (Kopelman, 1976), these highlighted naphthalene moieties may represent segments of alternative photochemistry strategies used by these species. Also, we showed high levels of 3MTPA in diatoms. As 3MTPA is toxic in high quantities to many organisms, diatoms may employ 3MTPA as a defense against grazers or competitors for nutrients in the vicinity of the species. DMSP, a biochemically related sulfur metabolite, has also been shown to act as a defense against grazers in addition to its antioxidant role (Sunda et al., 2002).

In conclusion, the new genomes presented in this study raise the number of publicly available microalgal genomes from less than 40 to greater than 60. We described findings from these new genomes as they pertain to organisms from new clades and saline environments compared with freshwater environments. The genomic and metabolomics datasets presented in this study significantly expand our knowledge, not solely regarding subtropical microalgae, but microalgae in general.

Limitations of the Study

Our study encompasses the microalgal species that we could gather in 5 years of bioprospecting, single-colony isolation, and sequencing. We are limited in our sampling sites and the number of species recovered, and we note that future efforts toward more comprehensive environmental surveys will be key to discover new coastal, subtropical species not included in this manuscript. We are also limited by the non-use of expressed mRNA for gene and functional predictions. Genome-only based approaches have the advantage of being more comprehensive, whereas examining expressed mRNA can reveal other information not included in this article, including other splice sites and transcript isoforms. Finally, our functional analyses regarding protein functions have not yet been experimentally verified in several of the species we sequenced, and they are only predictions based on high-confidence HMM alignments.

Data Description

We isolated several species of microalgae from subtropical, coastal environments in the UAE. Among these species, 25 single-colony isolates grew well enough to be cultured axenically to harvest DNA for genomic sequencing. Species of *Halamphora*, *Dunaliella*, *Nannochloris*, and *Chloroidium* comprised the majority of our isolates. We present their genomic assemblies in conjunction with other new genomic assemblies in this report as well as the metabolomic characterization of 14 of the isolates to support hypotheses on how halophytes algae can thrive in an environment rife with abiotic stressors. Eight major datasets are

included in this study, and their descriptions and digital object identifiers can be found in the Supplemental Information.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, two figures, four tables, can be found with this article online at https://doi.org/10.1016/j.isci.2018.12.035. The 8 accompanying datasets can be found at Mendeley at https://doi.org/10.17623/7k9tvk9zz.1.

ACKNOWLEDGMENTS

We thank the NYUAD High-Performance Computing and NYUAD Bioinformatics Core for assisting in sequence read processing; we thank NYUAD Core Technology Platforms for assistance in metabolomics studies and high-throughput sequencing. We thank John Burt, Grace Vaughan, Basel Khraiwesh, Joseph Koussa, Sujirat Yutap, and Jacobo Reyes-Velasco for providing some of the environmental samples. We thank Stephane Boissinot and Somayeh Ghorbani for discussions. New York University Abu Dhabi (NYUAD) Institute Grant (G1205-1205i) and NYUAD Faculty Research Funds (AD060) provided financial support for this work. Finally, we thank Sujirat Yutap for her work in designing the graphical abstract.

AUTHOR CONTRIBUTIONS

D.R.N. isolated strains, extracted DNA, and performed bioinformatic analyses. A.J., Z.H., K.M.H., and A.C. performed bioinformatic analyses. S.D. and W.F. maintained cultures of the featured strains and tested their heat tolerance. A.M. and S.D. carried out data mining for biogeographical description of the species. M.S. and M.A. performed library creation and sequencing. A.J. deposited the genomic data at NCBI. K.S.-A. and D.R.N. conceived the study, co-designed the experiments, and co-wrote the manuscript; all authors contributed to the editing of the manuscript. K.S.-A. supervised the execution of the study.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: July 24, 2018
Revised: November 7, 2018
Accepted: December 28, 2018
Published: January 25, 2019

REFERENCES

Alborzi, S.Z., Devignes, M.D., and Ritchie, D.W. (2017). ECDomainMiner: discovering hidden associations between enzyme commission numbers and Pfam domains. BMC Bioinformatics *18*, 107.

Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D.S., and Vardi, A. (2015). MARINE SULFUR CYCLE. Identification of the algal dimethyl sulfide-releasing enzyme: a missing link in the marine sulfur cycle. Science *348*, 1466–1469.

Almahasheer, H., Serrano, O., Duarte, C.M., Arias-Ortiz, A., Masque, P., and Irigoien, X. (2017). Low carbon sink capacity of Red Sea mangroves. Sci. Rep. *7*, 9700.

Arnold, H.E., Kerrison, P., and Steinke, M. (2013). Interacting effects of ocean acidification and warming on growth and DMS-production in the haptophyte coccolithophore *Emiliania huxleyi*. Glob. Chang. Biol. *19*, 1007–1016.

Bach, L.T., Alvarez-Fernandez, S., Hornick, T., Stuhr, A., and Riebesell, U. (2017). Simulated ocean acidification reveals winners and losers in coastal phytoplankton. PLoS One *12*, e0188198.

Balogh, L.M., and Atkins, W.M. (2011). Interactions of glutathione transferases with 4-hydroxynonenal. Drug Metab. Rev. *43*, 165–178.

Battin, E.E., and Brumaghim, J.L. (2009). Antioxidant activity of sulfur and selenium: a review of reactive oxygen species scavenging, glutathione peroxidase, and metal-binding antioxidant mechanisms. Cell Biochem. Biophys. *55*, 1–23.

Bose, H., and Satyanarayana, T. (2017). Microbial carbonic anhydrases in biomimetic carbon sequestration for mitigating global warming: prospects and perspectives. Front. Microbiol. *8*, 1615.

Brose, J., La Fontaine, S., Wedd, A.G., and Xiao, Z. (2014). Redox sulfur chemistry of the copper chaperone Atox1 is regulated by the enzyme glutaredoxin 1, the reduction potential of the glutathione couple GSSG/2GSH and the availability of Cu(I). Metallomics *6*, 793–808.

Cerda, M.M., and Pluth, M.D. (2018). S marks the spot: linking the antioxidant activity of N-acetyl cysteine to H2S and sulfane sulfur species. Cell Chem. Biol. *25*, 353–355.

Chaiboonchoe, A., Ghamsari, L., Dohai, B., Ng, P., Khraiwesh, B., Jaiswal, A., and Salehi-Ashtiani, K. (2016). Systems level analysis of the *Chlamydomonas reinhardtii* metabolic network reveals variability in evolutionary co-conservation. Mol. Biosyst. *12*, 2394–2407.

Chatzikonstantinou, M., Vlachakis, D., Chronopoulou, E., Papageorgiou, L., Papageorgiou, A.C., and Labrou, N.E. (2017). The glutathione transferase family of *Chlamydomonas reinhardtii*: identification and characterization of novel sigma class-like enzymes. Algal Res. *24*, 237–250.

Cooke, H.A., Guenther, E.L., Luo, Y., Shen, B., and Bruner, S.D. (2009). Molecular basis of substrate promiscuity for the SAM-dependent O-methyltransferase NcsB1, involved in the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin. Biochemistry *48*, 9590–9598.

Curson, A.R.J., Williams, B.T., Pinchbeck, B.J., Sims, L.P., Martinez, A.B., Rivera, P.P.L., Todd, J.D., Mercadé, E., Spurgin, L.G., and Carrión, O. (2018). DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. Nat. Microbiol. *3*, 430–439.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., and Karsenti, E. (2015). Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. Science *348*, 1261605.

Dixit, P., Mukherjee, P.K., Ramachandran, V., and Eapen, S. (2011). Glutathione transferase from *Trichoderma virens* enhances cadmium tolerance without enhancing its accumulation in transgenic *Nicotiana tabacum*. PLoS One *6*, e16360.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Ezcurra, P., Ezcurra, E., Garcillan, P.P., Costa, M.T., and Aburto-Oropeza, O. (2016). Coastal landforms and accumulation of mangrove peat increase carbon sequestration and storage. Proc. Natl. Acad. Sci. U S A *113*, 4404–4409.

Ezerina, D., Takano, Y., Hanaoka, K., Urano, Y., and Dick, T.P. (2018). N-acetyl cysteine functions as a fast-acting antioxidant by triggering intracellular H2S and sulfane sulfur production. Cell Chem. Biol. *25*, 447–459.e4.

Fassioli, F., Dinshaw, R., Arpin, P.C., and Scholes, G.D. (2014). Photosynthetic light harvesting: excitons and coherence. J. R. Soc. Interfaces *11*, 20130901.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., and Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279–D285.

Giordano, M., Norici, A., and Hell, R. (2005). Sulfur and phytoplankton: acquisition, metabolism and impact on the environment. New Phytol. *166*, 371–382.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.

Hader, D.P., Villafane, V.E., and Helbling, E.W. (2014). Productivity of aquatic primary producers under global climate change. Photochem. Photobiol. Sci. *13*, 1370–1392.

Haworth, M., Catola, S., Marino, G., Brunetti, C., Michelozzi, M., Riggi, E., Avola, G., Cosentino, S.L., Loreto, F., and Centritto, M. (2017). Moderate drought stress induces increased foliar dimethylsulphoniopropionate (DMSP) concentration and isoprene emission in two contrasting ecotypes of arundo donax. Front. Plant Sci. *8*, 1016.

Hell, R.D. (2008). Sulfur Metabolism in Phototrophic Organisms (Springer).

Hurtgen, M.T. (2012). Geochemistry. The marine sulfur cycle, revisited. Science *337*, 305–306.

Ibrahim, S., and Mohamed, G.A. (2016). Naturally occurring naphthalenes: chemistry, biosynthesis, structural elucidation, and biological activities. Phytochem. Rev. *15*, 279–295.

Kageyama, H., Tanaka, Y., Shibata, A., Waditee-Sirisattha, R., and Takabe, T. (2018). Dimethylsulfoniopropionate biosynthesis in a diatom *Thalassiosira pseudonana*: Identification of a gene encoding MTHB-methyltransferase. Arch. Biochem. Biophys. *645*, 100–106.

Kettles, N.L., Kopriva, S., and Malin, G. (2014). Insights into the regulation of DMSP synthesis in the diatom *Thalassiosira pseudonana* through APR activity, proteomics and gene expression analyses on cells acclimating to changes in salinity, light and nitrogen. PLoS One *9*, e94795.

Khan, M.I.R., Asgher, M., Iqbal, N., and Khan, N.A. (2013). Potentiality of Sulphur-Containing Compounds in Salt Stress Tolerance Ecophysiology and Responses of Plants under Salt Stress (Springer), pp. 443–472.

Kim, J., Hannibal, L., Gherasim, C., Jacobsen, D.W., and Banerjee, R. (2009). A human vitamin B12 trafficking protein uses glutathione transferase activity for processing alkylcobalamins. J. Biol. Chem. *284*, 33418–33424.

Kopelman, R. (1976). Exciton percolation in mixed molecular crystals and aggregates: from naphthalene to photosynthesis. J. Phys. Chem. *80*, 2191–2195.

Kumar, S., and Trivedi, P.K. (2018). Glutathione S-transferases: role in combating abiotic stresses including arsenic detoxification in plants. Front. Plant Sci. *9*, 751.

Lei, L., Cherukuri, K.P., Alcolombri, U., Meltzer, D., and Tawfik, D.S. (2018). The dimethylsulfoniopropionate (DMSP) lyase and lyase-like cupin family consists of bona fide DMSP lyases as well as other enzymes with unknown function. Biochemistry *57*, 3364–3377.

Lewin, J.C. (1954). Silicon metabolism in diatoms. I. Evidence for the role of reduced sulfur compounds in silicon utilization. J. Gen. Physiol. *37*, 589–599.

Li, Z.G., Ding, X.J., and Du, P.F. (2013). Hydrogen sulfide donor sodium hydrosulfide-improved heat tolerance in maize and involvement of proline. J. Plant Physiol. *170*, 741–747.

Maclean, I.M., and Wilson, R.J. (2011). Recent ecological responses to climate change support predictions of high extinction risk. Proc. Natl. Acad. Sci. U S A *108*, 12337–12342.

Martin, J.L., and McMillan, F.M. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr. Opin. Struct. Biol. *12*, 783–793.

Mukwevho, E., Ferreira, Z., and Ayeleso, A. (2014). Potential role of sulfur-containing antioxidant systems in highly oxidative environments. Molecules *19*, 19376–19389.

Nelson, D.R., Khraiwesh, B., Fu, W., Alseekh, S., Jaiswal, A., Chaiboonchoe, A., Hazzouri, K.M., O'Connor, M.J., Butterfoss, G.L., and Salehi-Ashtiani, K. (2017). The genome and phenome of the green alga *Chloroidium* sp. UTEX 3007 reveal adaptive traits for desert acclimatization. Elife *6*, https://doi.org/10.7554/eLife.25783.

Paerl, H.W., Gardner, W.S., Havens, K.E., Joyner, A.R., McCarthy, M.J., Newell, S.E., Qin, B., and Scott, J.T. (2016). Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients. Harmful Algae *54*, 213–222.

Pjevac, P., Kamyshny, A., Jr., Dyksma, S., and Mussmann, M. (2014). Microbial consumption of zero-valence sulfur in marine benthic habitats. Environ. Microbiol. *16*, 3416–3430.

Porollo, A. (2014). EC2KEGG: a command line tool for comparison of metabolic pathways. Source Code Biol. Med. *9*, 19.

Prade, L., Huber, R., Manoharan, T.H., Fahl, W.E., and Reuter, W. (1997). Structures of class pi glutathione S-transferase from human placenta in complex with substrate, transition-state analogue and inhibitor. Structure *5*, 1287–1295.

Raina, J.B., Tapiolas, D.M., Foret, S., Lutz, A., Abrego, D., Ceh, J., Seneca, F.O., Clode, P.L., Bourne, D.G., and Motti, C.A. (2013). DMSP biosynthesis by an animal and its role in coral thermal stress response. Nature *502*, 677–680.

Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Prlic, A., and Rose, P.W. (2018). NGL Viewer: web-based molecular graphics for large complexes. Bioinformatics *34*, 3755–3758.

Schafer, H., Myronova, N., and Boden, R. (2010). Microbial degradation of dimethylsulphide and related C1-sulphur compounds: organisms and pathways controlling fluxes of sulphur in the biosphere. J. Exp. Bot. *61*, 315–334.

Seneviratne, G. (2003). Global warming and terrestrial carbon sequestration. J. Biosci. *28*, 653–655.

Siddiqui, M.N., Mostofa, M.G., Akter, M.M., Srivastava, A.K., Sayed, M.A., Hasan, M.S., and Tran, L.P. (2017). Impact of salt-induced toxicity on growth and yield-potential of local wheat cultivars: oxidative stress and ion toxicity are among the major determinants of salt-tolerant capacity. Chemosphere *187*, 385–394.

Steele, R. (1978). Identification of 3-methylthiopropionic acid as an intermediate in mammalian methionine metabolism in vitro. J. Biol. Chem. *253*, 7844–7850.

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.

Tragin, M., and Vaulot, D. (2018). Green microalgae in marine coastal waters: the Ocean Sampling Day (OSD) dataset. Sci. Rep. *8*, 14020.

Wasmund, K., Mussmann, M., and Loy, A. (2017). The life sulfuric: microbial ecology of sulfur cycling in marine sediments. Environ. Microbiol. Rep. *9*, 323–344.

Wilson, R.A., Sangha, M.K., Banga, S.S., Atwal, A.K., and Gupta, S. (2014). Heat stress tolerance in relation to oxidative stress and antioxidants in *Brassica juncea*. J. Environ. Biol. *35*, 383–387.

Xu, J., Xing, X.J., Tian, Y.S., Peng, R.H., Xue, Y., Zhao, W., and Yao, Q.H. (2015). Transgenic Arabidopsis plants expressing tomato glutathione S-transferase showed enhanced resistance to salt and drought stress. PLoS One *10*, e0136960.

Yang, Z., Cheng, B., Xu, Y., Liu, D., Ma, J., and Ji, D. (2018). Stable isotopes in water indicate sources of nutrients that drive algal blooms in the tributary bay of a subtropical reservoir. Sci. Total Environ. *634*, 205–213.

Young, I.S., and McEneny, J. (2001). Lipoprotein oxidation and atherosclerosis. Biochem. Soc. Trans. *29* (Pt 2), 358–362.

ISCI, Volume *11*

Supplemental Information

Potential for Heightened Sulfur-Metabolic Capacity in Coastal Subtropical Microalgae

David R. Nelson, Amphun Chaiboonchoe, Weiqi Fu, Khaled M. Hazzouri, Ziyuan Huang, Ashish Jaiswal, Sarah Daakour, Alexandra Mystikou, Marc Arnoux, Mehar Sultana, and Kourosh Salehi-Ashtiani

SUPPLEMENTAL FIGURES

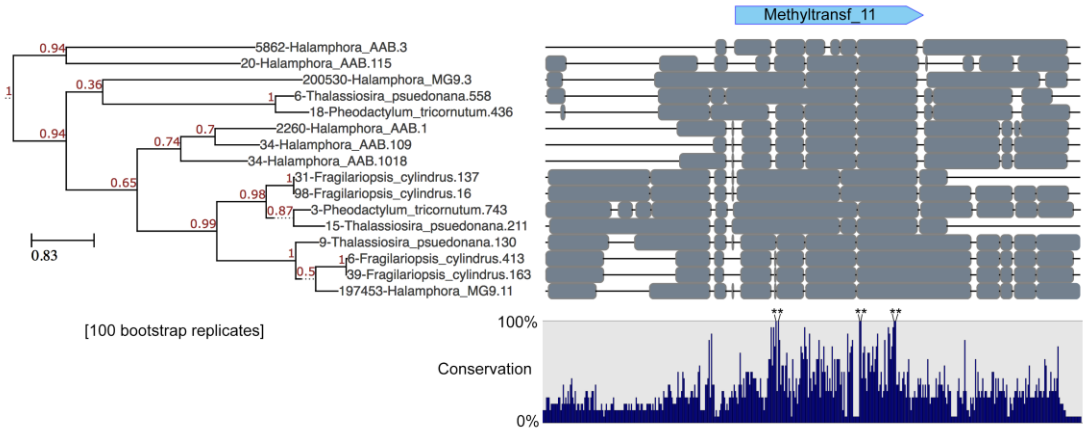


Fig. S1: Maximum likelihood phylogeny of aligned MTHB-MT genes. Related to Fig. 6. The tree was constructed using the ETE3 'standard_trimmed_bootstrap_raxml' workflow (-w) (Huerta-Cepas et al., 2016; Stamatakis, 2014, 2015). Bootstrap values are indicated at branch points. The region in each protein sequence corresponding to the HMM3-predicted methyltransf_11 PFAM domain, with E-values reported in Supplementary Table 1.

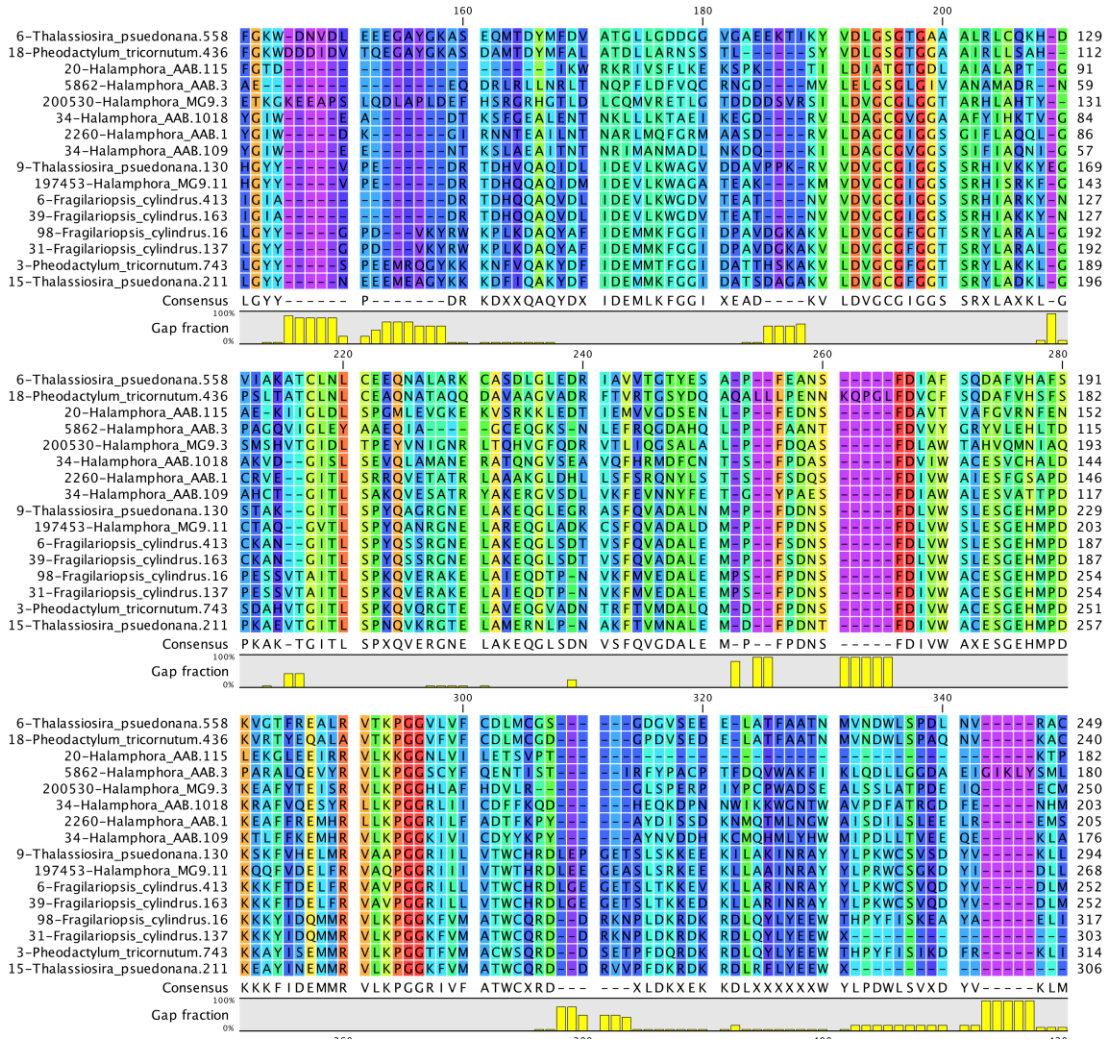


Fig. S2: Alignment of putative methylthiohdyroxybutyrate methyltransferases (MTHB-MTs) recovered from sequenced diatom species showing overall conservation and gap fractions. Related to Fig. 6.

SUPPLEMENTARY TABLES

Sequence	Domain	Start	End	Accession	Score	E-value
6-Thalassiosira_psuedonana.558	Methyltransf_11	111	211	PF08241.9	59.8	2.7E-16
5862-Halamphora_AAB.3	Methyltransf_11	42	135	PF08241.9	85.9	2E-24
200530-Halamphora_MG9.3	Methyltransf_11	114	213	PF08241.9	74.6	6.5E-21
34-Halamphora_AAB.1018	Methyltransf_11	66	164	PF08241.9	79.1	2.6E-22
2260-Halamphora_AAB.1	Methyltransf_11	68	166	PF08241.9	64.5	9.8E-18
34-Halamphora_AAB.109	Methyltransf_11	39	137	PF08241.9	64.9	7.3E-18
9-Thalassiosira_psuedonana.130	Methyltransf_11	150	249	PF08241.9	69.3	3E-19
197453-Halamphora_MG9.11	Methyltransf_11	125	223	PF08241.9	76.1	2.3E-21
6-Fragilariopsis_cylindrus.413	Methyltransf_11	109	207	PF08241.9	70.8	1E-19
39-Fragilariopsis_cylindrus.163	Methyltransf_11	109	207	PF08241.9	70.7	1.1E-19
98-Fragilariopsis_cylindrus.16	Methyltransf_11	174	274	PF08241.9	82.5	2.2E-23
31-Fragilariopsis_cylindrus.137	Methyltransf_11	174	274	PF08241.9	83.2	1.3E-23
3-Pheodactylum_tricornutum.743	Methyltransf_11	171	271	PF08241.9	83.7	9.8E-24
15-Thalassiosira_psuedonana.211	Methyltransf_11	178	277	PF08241.9	80	1.4E-22

Table S4: Description of the methyltransferase domains from the MTHB-MT candidates. Related to Fig. 6.

TRANSPARENT METHODS

Environmental sample collection: We obtained samples for microalgae isolation at locations designated by the approximated GPS coordinates in Additional file 1. Environmental material, either solid or liquid, was added to equal parts F/2 liquid growth media (Guillard's (F/2), (Lananan, Jusoh, Ali, Lam, & Endut, 2013)) and allowed to incubate for >48 hours under constant illumination (~100 µ mol photons/m^2) at 25°C. We then plated 20 µ l of the sample mixture on F/2 agar-based plates made with either filtered seawater (obtained from the Arabian Gulf at appx GPS coordinates: 24.593416, 54.347279) or MilliQ-purified (Merck-Millipore, Burlington, MA, USA) water and allowed to incubate for

two weeks. Initial isolation plates contained 200 µ g/ml Ampicillin (Anti-bacterial, Sigma-Aldrich, St. Louis, MO, USA), 50 µ g/ml Kanamycin (Anti-bacterial, Sigma-Aldrich, St. Louis, MO, USA), 200 µ g/ml Timentin (Anti-bacterial, mix of ticarcillin and sodium clavulanate, BioWorld, Dublin, OH, USA), and 20 µ g/ml Carbendazim (Anti-fungal, Sigma-Aldrich, St. Louis, MO, USA).

Microalgae isolation: We inspected plates for single colonies, and individual colonies were re-streaked to test for isogeneity. Single colony re-isolation was performed on each culture until we obtained axenic cultures. We tested new isolates for contamination by Gram Staining (Kaplan & Kaplan, 1933), phase contrast microscopy, and by re-streaking on rich media. We maintained cultures of isolated strains as we grew them at constant illumination (~120 µ mol photons/m^2) and 25°C until enough biomass was available to produce DNA for sequencing. More detailed protocols relating to the microbiology are in (Sharma et al., 2015).

Genomic sequencing: DNA was extracted from microalgae strains using Qiagen Plant Maxi Kits (Qiagen, Hilden, Germany). We used Illumina TruSeq PCR-free, Nextera mate-pair, and TruSeq DNA-nano kits for library creation (Illumina, San Diego, USA). Insert size ranged from 400-1200 bp for non-mate-pair libraries, and from 2000-4000, 5000-7000, and 8000-12000 bp for mate-pair libraries. We reconstructed whole-genome sequences for a variety of downstream comparative analyses including our functional annotations. We uploaded the genomes at NCBI under the Bioproject accession PRJNA428298.

Genome assembly and annotation: *De novo* assembly was performed using ABYSS (Jackman et al., 2017) with multiple k-mer size values. However, a k-mer length of 64 gave the best assembly for all tested; consequently, we used this value for every assembly reported here. Mate-pair libraries were used for scaffolding as indicated in Additional file 1. We uploaded whole-genome assemblies to NCBI under the submission ID: SUB3198369, Bioproject: PRJNA428298, Biosample: SAMN08335870, and Organism accessions QAWZ00000000 (*Nannochloris sp. X1*), SAMN08335869 (*Nannochloris sp. RS*), SAMN08335868 (*Halamphora sp. MG8b*), SAMN08335867 (Halamphora sp. AAB), SAMN08335866 (*Chlamydomonas sp. NG2*), SAMN08335865 (*Dunaliella sp. YS1*), SAMN08335864 (*Dunaliella sp. WIN1*), SAMN08335863 (*Dunaliella sp. RO*), SAMN08335861 (*Chloromonas sp. AAM2*), SAMN08335860 (*Chloroidium sp. JM*), SAMN08335859 (*Chloroidium sp. CF*), SAMN08335858 (*Chlorella sp. KRBP*), SAMN08335857 (*Chlamydomonas sp. WS7*), SAMN08335856 (*Chlamydomonas sp. WS3*), SAMN08335854 (*Chlamydomonas sp. 3222*), SAMN08335853 (*Chlamydomonas sp. 3112*), SAMN08335852 (*Characiochloris sp. AAM3*). In the case of *Chloroidium sp. CF*, this genome was updated from the previously published *Chloroidium sp. UTEX3077* (Nelson et al., 2017) with long-, mid-, and short-range mate-pair insert libraries. In the updated version of the assembly, we noticed that gene content, including functional annotations, did not change significantly from the previously published version (Nelson et al., 2017); however, the assembly scaffold N50 increased from 150kpb to 2.2Mbps.

Genome statistics were calculated using QUAST (Gurevich, Saveliev, Vyahhi, & Tesler, 2013); coding regions (exons) were predicted *de novo* using an *Arabidopsis thaliana* Hidden Markov Model (HMM) in SNAP (Korf, Flicek, Duan, & Brent, 2001) (Dataset 2). We performed protein family domain prediction using HMMER v3.1b2 (www.hmmer.org). The arguments for Hmmsearch were '--cpu 28 --noali --domtblout'. For comparisons using confidence score totals between two groups of species, such as those in Fig. 4, two-tailed t-tests were performed with the indicated p-values. The CLC Genomics Workbench (v10.0, https://www.qiagenbioinformatics.com) was used to create the protein models and Kyle-Doolittle hydrophobicity plots in Fig. 6.

Data obtained from public repositories: We downloaded the genomes from 45 photosynthetic eukaryotes, mostly microalgae, for use in our comparative genomics studies. Descriptions of each genome used, including links to the assembly download site, the organization behind the sequencing initiative, and any relevant publications are listed in the Excel file in Dataset 3.

Phylogeny reconstructions: Ribulose-1,5-bisphosphate carboxylase/oxygenase genes were extracted from BlastP (Altschul, Gish, Miller, Myers, & Lipman, 1990) results (Dataset 1) and aligned using MUSCLE (Robert C. Edgar (https://www.ebi.ac.uk/Tools/msa/muscle/, v3.8.31). We then used these alignments to perform evolutionary model testing using the ETE 3 Python package (Huerta-Cepas, Serra, & Bork, 2016) (http://etetoolkit.org/). Other software included in this package, used in this manuscript, but not listed in the figure legends include: version 2.1.8, Double precision (No SSE3), kalign, Kalign version 2.04, Copyright 2004, 2005, 2006 Timo Lassmann, mafft: MAFFT v6.861b (https://mafft.cbrc.jp, 2011/09/24), MUSCLE by Robert C. Edgar (https://www.ebi.ac.uk/Tools/msa/muscle/, v3.8.31), phyml (PhyML version 20160115), RAxML v8.1.20 by Alexandros Stamatakis (https://github.com/stamatak/standard-RAxML, v8.1.20), T-COFFEE (tcoffee.crg.cat, v11.00.8cbe486 (2014-08-12 22:05:29 - Revision 8cbe486 - Build 477)), and trimAl (trimal.cgenomics.org/, v1.4.rev6

build[2012-02-02]). All plots were made using publically available Python Plot.ly API libraries (https://plot.ly/api/).

The sequence for the methylthiohydroxybutyrate methyltransferase from *Thalassiosira pseudonana* involved in DMSP biosynthesis (MTHB-MT) was downloaded from NCBI and is described in (Kettles, Kopriva, & Malin, 2014). We performed BlastP searches to identify homologs of the MTHB-MT on the sequence in our annotations for *T. psuedonana* and retrieved one or more sequences as candidate MTHB-MT genes from the following diatoms with an e-value of 10e-10 or lower: *Thalassiosira pseudonana* (self-hit), *Phaeodactylum tricornutum*, *Fragilariopsis cylindrus*, *Halamphora sp. MG9 (*new isolate*),* and *Halamphora sp. AAB* (new isolate). For the final alignment and tree construction (Supp. Figs 1 and 2), only sequences with HMM matches of E-value 1e-16 or lower were used. The methyltransferase protein family domain model (Methyltransf_11) currently in the PfamScan database (https://www.ebi.ac.uk/Tools/pfa/pfamscan/) is 95 amino acids long, and this is the region of the gene with shared homology between all of the species shown in Fig. 5. The alignment region of the TpMMT gene with respect to the Methytransf_11 HMM is: 'fDlvvssevlhhv..edpekalkeiaRvLkpgGIlvi+Dl+vs+e++ hv ++e++l+e++RvL+pgG+l++ IDLCVSTEAFLHVgpGNHEAVLREAWRVLRPGGRLIF', where capital letters represent conserved residues, lowercase letters represent residues in an insertion with dots (".") as positional padding, and "+" symbols indicate conservative amino acid substitutions thus a positive score. The methytransf_11 domain is annotated as an S-adenosyl-L-methionine dependent methyltransferase and forms the Rossman-like alpha-beta fold typical of SAM-dependent methyltransferases (Zhang & Cheng, 2003) (Das et al., 2004) (Keller et al., 2002). The recovered sequences were aligned using MUSCLE (MUltiple SequenCe alignment by Log Expectation) with the following parameters: Find Diagonals (YES), 100 maximum iterations, 100,000 maximum hours, and 100 GB maximum memory usage (Additional File 4).

Metabolomics using LCMS-QToF: All species were inoculated on F/2 medium-containing plates (for freshwater species, distilled water was used instead of seawater) and incubated at 25 °C for one month. Then, lawns were collected, pelleted, and underwent microwave-assisted methanol extraction preceded by 30 minutes of sonication (adapted from (Annegowda, Bhat, Min-Tze, Karim, & Mansor, 2012)) and followed by filtration through a 0.2 µ m filter. We based the HPLC parameters on methods developed in (Hu, Wang, & Han, 2017) and (Nielsen, Maeda, & Bilgin, 2017). Briefly, compounds were eluted from a reverse-phase C18 column using a semi-linear isopropanol gradient. Eluted compounds went to an Agilent LC-MS QToF 6538 with accurate mass profiling (error <2ppm, Agilent, Santa Clara, CA, USA). We used XCMS (Benton et al., 2015; Tautenhahn, Patti, Rinehart, & Siuzdak, 2012) and the METLIN database (Smith et al., 2005; Tautenhahn, Cho, et al., 2012) for compound assignment and comparative statistics as well as in generating the TIC-cloud plot (Patti et al., 2013). Statistically significant differences in metabolite abundance were found using two-tailed t-tests, and we provide p-values adjusted for a false discovery rate (FDR) of 0.05.

SUPPLEMENTAL REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410. doi:10.1016/S0022-2836(05)80360-2

Annegowda, H. V., Bhat, R., Min-Tze, L., Karim, A. A., & Mansor, S. M. (2012). Influence of sonication treatments and extraction solvents on the phenolics and antioxidants in star fruits. *J Food Sci Technol, 49*(4), 510-514. doi:10.1007/s13197-011-0435-8

Benton, H. P., Ivanisevic, J., Mahieu, N. G., Kurczy, M. E., Johnson, C. H., Franco, L., . . . Siuzdak, G. (2015). Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal Chem, 87*(2), 884-891. doi:10.1021/ac5025649

Das, K., Acton, T., Chiang, Y., Shih, L., Arnold, E., & Montelione, G. T. (2004). Crystal structure of RlmAI: implications for understanding the 23S rRNA G745/G748-methylation at the macrolide antibiotic-binding site. *Proc Natl Acad Sci U S A, 101*(12), 4041-4046. doi:10.1073/pnas.0400189101

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072-1075. doi:10.1093/bioinformatics/btt086

Hu, C., Wang, M., & Han, X. (2017). Shotgun lipidomics in substantiating lipid peroxidation in redox biology: Methods and applications. *Redox Biol, 12*, 946-955. doi:10.1016/j.redox.2017.04.030

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol, 33*(6), 1635-1638. doi:10.1093/molbev/msw046

Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., . . . Birol, I. (2017). ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res, 27*(5), 768-777. doi:10.1101/gr.214346.116

Kaplan, M. L., & Kaplan, L. (1933). The Gram Stain and Differential Staining. *J Bacteriol, 25*(3), 309-321.

Keller, J. P., Smith, P. M., Benach, J., Christendat, D., deTitta, G. T., & Hunt, J. F. (2002). The crystal structure of MT0146/CbiT suggests that the putative precorrin-8w decarboxylase is a methyltransferase. *Structure, 10*(11), 1475-1487.

Kettles, N. L., Kopriva, S., & Malin, G. (2014). Insights into the regulation of DMSP synthesis in the diatom Thalassiosira pseudonana through APR activity, proteomics and gene expression analyses on cells acclimating to changes in salinity, light and nitrogen. *PLoS One, 9*(4), e94795. doi:10.1371/journal.pone.0094795

Korf, I., Flicek, P., Duan, D., & Brent, M. R. (2001). Integrating genomic homology into gene structure prediction. *Bioinformatics, 17 Suppl 1*, S140-148.

Lananan, F., Jusoh, A., Ali, N., Lam, S. S., & Endut, A. (2013). Effect of Conway Medium and f/2 Medium on the growth of six genera of South China Sea marine microalgae. *Bioresour Technol, 141*, 75-82. doi:10.1016/j.biortech.2013.03.006

Nelson, D. R., Khraiwesh, B., Fu, W., Alseekh, S., Jaiswal, A., Chaiboonchoe, A., . . . Salehi-Ashtiani, K. (2017). The genome and phenome of the green alga Chloroidium sp. UTEX 3007 reveal adaptive traits for desert acclimatization. *Elife, 6*. doi:10.7554/eLife.25783

Nielsen, I. O., Maeda, K., & Bilgin, M. (2017). Global Monitoring of the Mammalian Lipidome by Quantitative Shotgun Lipidomics. *Methods Mol Biol, 1609*, 123-139. doi:10.1007/978-1-4939-6996-8_12

Patti, G. J., Tautenhahn, R., Rinehart, D., Cho, K., Shriver, L. P., Manchester, M., . . . Siuzdak, G. (2013). A view from above: cloud plots to visualize global metabolomic data. *Anal Chem, 85*(2), 798-804. doi:10.1021/ac3029745

Sharma, S. K., Nelson, D. R., Abdrabu, R., Khraiwesh, B., Jijakli, K., Arnoux, M., . . . Salehi-Ashtiani, K. (2015). An integrative Raman microscopy-based workflow for rapid in situ analysis of microalgal lipid bodies. *Biotechnol Biofuels, 8*, 164. doi:10.1186/s13068-015-0349-1

Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., . . . Siuzdak, G. (2005). METLIN: a metabolite mass spectral database. *Ther Drug Monit, 27*(6), 747-751.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Stamatakis, A. (2015). Using RAxML to Infer Phylogenies. *Curr Protoc Bioinformatics, 51*, 6 14 11-14. doi:10.1002/0471250953.bi0614s51

Tautenhahn, R., Cho, K., Uritboonthai, W., Zhu, Z., Patti, G. J., & Siuzdak, G. (2012). An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat Biotechnol, 30*(9), 826-828. doi:10.1038/nbt.2348

Tautenhahn, R., Patti, G. J., Rinehart, D., & Siuzdak, G. (2012). XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem, 84*(11), 5035-5039. doi:10.1021/ac300698c

Zhang, X., & Cheng, X. (2003). Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. *Structure, 11*(5), 509-520.