EUKARYOTIC PHYTOPLANKTON COMMUNITY SPATIOTEMPORAL DYNAMICS AS IDENTIFIED THROUGH GENE EXPRESSION WITHIN A EUTROPHIC ESTUARY

Weida Gong

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Marine Sciences Department in the College of Arts and Sciences.

Chapel Hill
2017

Approved by:

Adrian Marchetti

Hans Paerl

Marc Alperin

Scott Gifford

ABSTRACT

Weida Gong: Eukaryotic phytoplankton community spatiotemporal dynamics as identified through gene expression within a eutrophic estuary
(Under the direction of Adrian Marchetti)

Estuaries are highly dynamic and productive environments. A clear understanding of how phytoplankton, supporting the base of the food web, respond to spatiotemporal dynamics is necessary to ensuring the health and sustainability of these ecosystems. Over the span of a year, we investigated the interactions between biotic and abiotic factors within the eutrophic Neuse River Estuary (NRE). Through metatranscriptomic sequencing in combination with water quality measurements, we show that there are different metabolic strategies deployed along the NRE, and nitrogen availability is the main driving factor for such divergence. In the upper estuary, phytoplankton express more transcripts of genes for synthesis of cellular components and carbon metabolism whereas in the lower estuary, transcripts allocated to nutrient metabolism and transport were more highly expressed. We advocate for the use of molecular sequencing approaches to complement coastal water quality monitoring programs as a way to examine microbial community dynamics in response to changing environmental conditions.

ACKNOWLEDGMENTS

I want to thank J. Browne, J. Braddy and other members of the ModMon Team who assisted with sample collection and analysis. I am grateful to H. Masters for assistance with RNA extractions and C. Stackhouse and J. Roach for assistance with sequence analysis. I also thank N. Cohen, C. Moreno and R. Lampe for their assistance in the development of the bioinformatics pipeline. UNC Research Computing generously provided cluster time for sequence analysis.

ModMon is funded by NC-Dept. of Environment and Natural Resources, Division of Water Resources, the Lower Neuse Basin Association and Neuse River Compliance Association and National Fish and Wildlife Foundation. This research was supported by NC Sea Grant projects R/12-HCE-3 - NA10OAR4170080 and R/14-HCE-2 – NA14OAR4170073 to A.M. and H.P. as well as start-up funds to A.M.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBRREVIATIONS

Chl *a*	chlorophyll *a*
Fe	iron
FeT	iron(III) transport system
KO	KEGG Orthology
MMETSP	Marine Microbial Eukaryote Transcriptome Sequencing Project
ModMon	Neuse River Estuary Modeling and Monitoring
MST	putative multiple sugar transporters
N-glycan	N-glycan biosynthesis
NO_3As	nitrate assimilation
NR	nitrate reduction
NRE	Neuse River Estuary
PCA	principal component analysis
PO_4T	phosphate transport system
PS	photosynthesis
PSR	phosphate starvation response system
SST	putative simple sugar transport system
UreaT	urea transport system
WGCNA	Weighted Gene Co-expression Network Analysis

INTRODUCTION

Estuaries are semi-enclosed water bodies that link freshwater to marine environments. Cross-estuary and along estuary gradients contribute to strong spatial differences in nutrient profiles (Reed et al. 2004). The highly dynamic hydrological and nutrient cycling properties result in a significant amount of biological variation. Steep spatial gradients often make the upper and lower sections of estuaries very different environments. Apart from spatial variation, temporal variation in precipitation, temperature, wind patterns, and turbidity also contribute to the dynamic nature of estuaries (Hall et al. 2012; Paerl et al. 2014; Day et al. 2012; Peierls et al. 2012). The temporal and spatial variations force biotic components of the ecosystem, including the critically-important phytoplankton which support the base of the estuarine food web, to vary in composition and physiology depending on the environmental conditions. In the Neuse River Estuary (NRE), phytoplankton decrease NO_3^- uptake along the NRE and depend primarily on recycled N forms such as ammonium/urea in the lower estuary (Twomey et al. 2005). In summer, phytoplankton increase ammonium utilization in the lower estuary due to the high seasonal stratification-induced remineralization (Twomey et al. 2005; Paerl et al. 1998). This hydrological variability is also reflected in the seasonal pattern of phytoplankton community composition. For example, dinoflagellates are usually a significant part of the community in spring while cyanobacteria can be dominant in summer (Paerl et al. 2010).

Phytoplankton communities in estuaries are directly affected by exogenous nutrient inputs. Increases in anthropogenic-derived nutrients in the past century have led to massive eutrophication and increased frequency and intensity of algal blooms in estuaries around the globe (Howarth et al. 1997; Nixon 1995). Increases in natural perturbations such as storms have also contributed to the sporadic nutrient loading in estuaries (Paerl 2006). In fact, anthropogenic nutrient inputs can sometimes be overshadowed by climatic events such as intense storms and hurricanes (Paerl et al. 2010; Paerl et al. 2014). Nutrient loading directly affects phytoplankton communities through increased phytoplankton growth, however, the relationship between nutrient loading and phytoplankton dynamics is complex and often non-linear due to the profound interplay between biological, chemical and physical properties throughout the water column (Peierls et al. 2012; Paerl et al. 2014).

Environmental conditions are quite different along the axis of the NRE. In the upper estuary, the "geochemical filter" caused by flocculation and aggregation of humic acids and metal ions sometimes can lead to light being a limiting factor for phytoplankton growth (Sharp et al. 1984). As this upper section is quite narrow, flow rates are typically high, especially in winter and early spring. Elevated riverine discharge increases nutrient loading and provides a favorable environment for phytoplankton growth. However, when discharge is too high, nutrient loading exceeds phytoplankton's assimilatory and growth rates and advective losses overwhelm production rates so that phytoplankton abundance decreases (Hall et al. 2012). Noticeably, these residual nutrients can be utilized during low flow periods and can sometimes contribute to high phytoplankton biomass in the lower estuary (Paerl et al. 1995; Paerl et al. 2014). As the NRE widens down estuary, water residence time increases. Phytoplankton's assimilatory capacity exceeds the flushing rate, and with sporadic nutrient loading, blooms frequently occur. The relationship between nutrient loading and the phytoplankton community is more complex if biotic components, such as grazing pressure, are also considered (Walz & Welker 1998; Brussaard 2003; Cloern 2001).

Phytoplankton community composition also varies spatiotemporally within the estuary as different phytoplankton groups can often display distinct responses to a similar set of environmental conditions. Diatoms, dinoflagellates, cryptophytes, chlorophytes and cyanobacteria are all represented in the NRE (Pinckney et al. 1998), and can display diverse responses to hydrographic conditions. Dinoflagellate blooms are common as their relatively large size and low susceptibility to grazing provide them with a competitive advantage. Their motility also plays an important role in winter/early spring when riverine discharge is high (Graham & Strom 2010; Demir et al. 2008; Walz & Welker 1998; Hall & Pearl 2011). Other than dinoflagellates, cryptophyte and chlorophyte blooms occur throughout the year, primarily coinciding with sporadic nutrient inputs (Hall et al. 2012). In summer months when temperature is high, phytoplankton that have high intrinsic growth rates at elevated temperatures, such as cyanobacteria and chlorophytes, can be abundant (Paerl et al. 1995). Diatoms have high maximum intrinsic growth rates (Smayda 1997) and are also a major component of the phytoplankton community in the NRE, although blooms rarely occur (Pinckney et al. 1998). Unlike dinoflagellates and cryptophytes, diatoms are only weakly linked to riverine inputs and are mostly affected by resuspension frequency (Hall et al. 2012).

Massive nutrient inputs and contemporaneous climatic events pose a severe threat to estuarine water quality and phytoplankton communities (Kennish & Paerl 2010). To understand how phytoplankton respond to these anthropogenic and natural perturbations, efforts have been made to study the correlation between phytoplankton communities and environmental parameters in the NRE (Paerl et al. 2014; Pinckney et al. 1997; Hall et al. 2012). Bulk plankton community analyses (e.g., chl *a*, primary productivity, nutrient uptake rates, etc.) and water quality measurements incorporated into models to study estuarine phytoplankton dynamics have been informative (Carstensen et al. 2007; Pinckney et al. 1997). However, the correlation can sometimes be masked due to the interplay among spatiotemporal dynamics in the system. For example, typically pulses of DIN cause the most significant increase in phytoplankton abundance in the NRE, however there can be periods of high phytoplankton biomass without preceding DIN pulses, and chl *a* can sometimes be negatively correlated with DIN (e.g., Peierls et al. 2012). Thus, such uncertainties in phytoplankton dynamics based of bulk measurements pose a serious challenge to understanding phytoplankton responses to environmental gradients. Knowledge on the physiology of taxonomic subsets within the plankton community can be informative to detecting distinct responses to environmental gradients and is needed to provide a more comprehensive view to better manage our valuable estuaries and coastal environments.

Metatranscriptomics has proven to be a powerful tool to provide insights into the inferred metabolic physiology of marine plankton within mixed natural assemblages (Cooper et al. 2014; Marchetti et al. 2012). Here, we have conducted a yearlong metatranscriptomic analysis of the plankton communities along a transect in the NRE. Results indicate phytoplankton communities displayed different response patterns to the steep nutrient gradients between the upper and lower estuary, and that different phytoplankton groups generally used similar strategies in such responses.

MATERIALS AND METHODS

Study area and sample collection.

Sampling for this study took place in 2012 in conjunction with the Neuse River Estuary Modeling and Monitoring (ModMon) Program (http://www.unc.edu/ims/neuse/modmon) run by the Institute of Marine Sciences, UNC-Chapel Hill and the North Carolina Department of Environmental and Natural Resources (now Deprtment of Environmental Quality) that has collected water samples in the NRE on a bimonthly or monthly basis since 1994. Metatranscriptomic sampling at four of the 11 routinely sampled stations (Modmon stations 20, 70, 120 and 180) were performed in conjunction to the field physical-chemical measurements collected as part of the ModMon program (Fig. 1B).

Environmental measurements.

Vertical profiles of temperature, salinity, dissolved oxygen, in vivo fluorescence and light were collected with a YSI 6000 multiprobe sonde coupled to a LiCor LI-1925A quantum sensor that records photosynthetically active radiation (PAR; 400-700 nm) (Yellow Springs, Inc., Yellow Springs, OH, USA). Soluble nutrients (including nitrite/nitrate [NO_2^- NO_3^-], phosphate [PO_4^{3-}] and silicate [SiO_4^{2-}]), were measured using a Lachat Quick-chem 8000 auto-analyzer (Lachat, Milwaukee) as described in Peierls et al. (2012). Chl *a* concentrations were determined using the non-acidification method of Welschmeyer (1994) on a Turner Designs Trilogy fluorometer as described in Peierls et al. (2012) and Welschmeyer (1994). Primary productivity was measured using the $NaH^{14}CO_3$ incorporation method as modified by

Mallin and Paerl (1992). Photopigment analysis via HPLC was performed as described by Pinckney et al. (1996) except that the analyses were performed on a Shimadzu LC-20AB HPLC coupled to a Shimad SPD M20A in-line photodiode array spectrophotometer (Shimadzu Inc. USA) (Pinckney et al. 1996).

RNA-Seq sample preparation and sequencing.

Metatranscriptomic sampling occurred in the months of February, April, June, August and December of 2012 to encompass the seasonal succession of the phytoplankton assemblages in the NRE. Near surface (~0.2 m depth) samples for RNA were collected and immediately filtered onto multiple Millipore isopore membrane filters (0.45 μm, 142 mm) per station directly using a 3-head Masterflex L/S peristaltic pump with pre-rinsed Tygon tubing and screened from direct sunlight. Filters were changed every 15 minutes or when the flow of water decreased due to particle clogging. Individual filters were placed in Ziploc bags, wrapped in aluminum foil, and immediately placed in liquid nitrogen. A minimum of six filters were collected from each station and sampling date. Based on pump flow rates, an estimated volume of 25 L was filtered per station. Onshore, filters were stored at -80°C until RNA extractions were performed (typically within one month).

For RNA extractions, filters were briefly thawed on ice. RNA was extracted from individual cut-up filters using the ToTALLY RNA Kit (Ambion) according to the manufacturer's protocols with the additional step that filter pieces were first vortexed in 7 mL of denaturation solution containing 0.5 mL of glass beads and the resulting lysate was centrifuged at 8,801 x *g* and 4°C for 3 min. Trace DNA contamination was reduced by DNase 1 (Ambion) digestion at 37°C for 45 min. Polyadenosine [poly(A)+] RNA (mRNA) was isolated with the

MicroPoly(A) Purist Kit (Ambion) according to the manufacturer's instructions. mRNA samples from several filters at each station were then combined to achieve a minimum total of 100 ng of mRNA. Illumina sequence library synthesis using the TruSeq mRNA Library Preparation Kit (San Diego, CA, USA) was performed at the UNC High-throughput Sequencing Facility. Samples of pooled mRNA from each time point were bar-coded and sequenced on a single lane of the Illumina HiSeq2000 platform (San Diego, CA, USA), generating between 53–123 million 100 bp paired-end reads per sample.

Sequence assembly, taxonomic identification and functional gene annotation pipeline.

FastQC was used to assess read quality (Andrews 2010). Paired-end sequence reads from each sample were individually assembled into larger transcripts (termed contigs) using ABySS v 1.3.5 with multiple k-mer sizes (from 52 to 96 with a step of 2). Using Trans-ABySS v1.4.4, contigs were filtered and merged, and reads were mapped to the merged contigs (Simpson et al. 2009; Robertson et al. 2010). The number of sequence reads that aligned to each contig was calculated with SAMtools v0.1.19, and tabulated for differential gene expression analysis using the Caroline package in R (Schruth 2013; Li et al. 2009). MarineRefII (http://ssharma.marsci.uga.edu/Lab/MarineRef2/), a custom-made reference organism database of marine microbial eukaryotes and prokaryotes (maintained by the Moran Lab at the University of Georgia, Athens) was used for taxonomical annotation that includes all sequenced transcriptomes that are part of the Marine Microeukaryote Transcriptome Project (MMETSP) (Keeling et al. 2014). A sequence similarity search of the assembled contigs against MarineRefII was performed through BLASTx (v. 2.2.28), with an *e*-value cutoff of 0.001. Only reference contigs with the lowest *e*-values were kept as best hits for subsequent analysis. Each hit's taxonomic ID was used as an entry to obtain taxonomic information from the NCBI Taxonomy Database. Due to discrepancies between the NCBI database and widely-used phytoplankton taxonomic ranks, a manually curated taxonomic table was used to provide consistent taxonomic classifications for organisms with best hits to the assembled contigs (Appendix 2) (Keeling et al. 2014).

For functional gene annotations, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used as a reference database. An *e*-value cutoff of 0.001 was similarly applied for KEGG homology searches, where the sequence with the lowest *e*-value was assigned. If this sequence lacked a KEGG module annotation, the next-best sequence reference contigs were screened until a sequence with a KEGG module annotation was identified. Read counts of contigs with an assigned KEGG Orthology (KO) identity were used for differential expression analysis. Best hit results from the taxonomic and functional gene annotation searches were then merged for each contig, along with the quantitative read count information for each sample. Assembled contigs and normalized counts at the module level for all samples were also archived (Dataset 1 and 2; http://marchettilab.web.unc.edu/data/). This final composite dataset includes a list of contigs with associated functional gene and taxonomical annotations as well as quantitative read counts (Dataset 3; http://marchettilab.web.unc.edu/data/).

Differential Expression analysis.

Samples were clustered based on sampling site (see results), and KOs were normalized using "trimmed means of M (TMM)" normalization procedure available in the edgeR package. The TMM method normalizes reads by computing a scaling factor after excluding genes that have high average counts and/or have large expression differences between samples with the assumption that most of the genes are not differentially expressed (Robinson & Oshlack 2010). Principle component analyses (PCA) were used with environmental measurement data collected over a 5-year period (2007 – 2012) and TMM-normalized KO data (20 metatranscriptomic samples) to visualize the spatiotemporal variability. EdgeR was used to detect differentially expressed KOs between Station 20 (the upper station) and the lower stations by way of the MANTA package as described in Marchetti et al. (2012), and significance of differentially expressed KOs were assigned with the 'exactTest' program (Robinson et al. 2009). Normalized reads for each KO were then grouped into modules using their KO associated "Module ID". Only reads with KEGG module annotations were used for differential expression analysis. A Mann-Whitney U (MWU) test was conducted to determine which modules were significantly enriched (Nielsen et al. 2005). Modules that displayed a significant difference (p-value < 0.05) in the combined read count ratio between stations relative to the TMM were considered either over-represented (above TMM) or under-represented (below TMM) (Appendix 3).

Differential expression analysis was also performed on each of the four dominant eukaryotic phytoplankton functional groups (i.e., diatoms, dinoflagellates, chlorophytes and cryptophytes). KOs were grouped at the KEGG clas3 level. Phytoplankton's responses were represented by the fold change of transcripts at clas3 level, and fold change values were then used to calculate the variances among the different phytoplankton groups. KEGG clas3 gene groupings with high variances within the upper quarter were defined as metabolic functions with different expression patterns among various phytoplankton groups. KEGG clas3 groupings with lower variances are considered as similarly expressed among groups.

Correlation and network analyses.

To assess the correlations between gene expression patterns and environmental measurements, normalized counts of KOs were subjected to a Weighted Gene Correlation Network Analysis (WGCNA) (Langfelder & Horvath 2008). WGCNA clusters KOs into modules (ME, designated by colors, hereafter denoted as subnetworks to avoid confusion with KEGG modules) based on their expression dissimilarity and demonstrates the correlation between each subnetwork and environmental measurements. Fisher-exact test of the binary measure of a KO's presence was then applied to identify which metabolic functions (KEGG clas3) were enriched within each subnetwork. Correlation analysis was also conducted between KEGG modules and environmental measurements using Pearson correlation with associated *p*-values assigned to construct a similarity matrix. The similarity matrix was then used to build a network map, for which each node is a KEGG module or environmental measurement and each edge between nodes represents statistically significant correlations. Network maps were created with the R package igraph (v 1.0.0) and graphs were generated using Cytoscape (v 3.3.0, www.cytoscape.org).

RESULTS

Spatiotemporal dynamics in the NRE

The NRE is a major tributary of North Carolina's Albemarle-Pamlico Estuarine System, the USA's largest lagoonal estuarine system (Fig. 1A). Characterized by significant seasonal/temporal and spatial variations of hydrologic processes and environmental factors, the NRE constitutes a highly dynamic ecosystem for all organisms, including phytoplankton. Salinity gradually increases from the head to the mouth of the estuary (Fig. 1B, Table 1). In contrast to the salinity profile, as high nutrient riverine runoff is diluted by low nutrient sea water along with nutrient utilization by phytoplankton, nutrient concentrations (e.g., NO_3^- and PO_4^{3-}) decrease dramatically moving downstream in the estuary (Fig. 1B, Table 1). In addition to spatial gradients, the NRE also displayed high seasonal/inter-annual variability, which is reflected by high standard deviations of nutrient concentrations particularly in the upper and mid-estuary (Fig. 1B). The high variability of environmental factors and interplay of biological and hydrological processes weakened the relationship with phytoplankton biomass – none of these environmental measurements (apart from POC, which is another indicator of biomass) had strong correlations with chl *a* (Fig. 1C). We applied a PCA to all the samples collected between 2007 and 2012 to examine how environmental conditions differed spatiotemporally. Clusters that are representative of the lower estuary stations (station 70, 120, 180) largely overlap with each other, indicative of similar environmental conditions during our sampling period (Fig. 1D). The cluster for station 20 is separated from the lower estuary stations, with nutrient concentrations being the major driver of the observed divergence (Fig. 1D). The first principle component (PC1) is mostly represented by the horizontal components of NO_3^-/NO_2^-, NH_4^+ and salinity profiles, which are reflective of spatial variation in the NRE. The second principle component (PC2) is mainly represented by the vertical components of dissolved oxygen (DO) and temperature, which are reflective of temporal variations in the NRE (Fig. 1D). Nutrient-driven spatial gradients explain 41.7% of the variance among samples while temperature/DO-driven temporal gradients explain 27.5%, indicating that spatial differences are more significant than temporal ones.

Metatranscriptome sequencing statistics.

Numbers of paired-end raw reads sequenced within each of the metatranscriptomic samples used in this study ranged between 53-123 million per sample (Table 2). Assembly of reads yielded 0.4-1.2 million contigs for each sample. Between 47-63% and 41-56% of raw reads had sequence similarity hits to the reference sequences in the taxonomic annotation database, MarineRefII, and functional annotation database, KEGG, respectively. Between 7-11% of raw reads in each sequence library were mapped to KEGG reference sequences containing at least one KEGG module annotation (Table 2).

Taxonomic annotation and patterns of gene expression in the NRE.

Diatoms, dinoflagellates, chlorophytes and cryptophytes are typically the most abundant eukaryotic phytoplankton groups, each commonly contributing equally (around 20%) to the total phytoplankton chl *a* in the NRE (Pinckney et al. 1998). Coincidently, more transcripts were allocated to these four groups than any others, although diatom and dinoflagellate-associated transcripts were more abundant than chlorophytes and cryptophytes (Fig. 2A). Analysis of

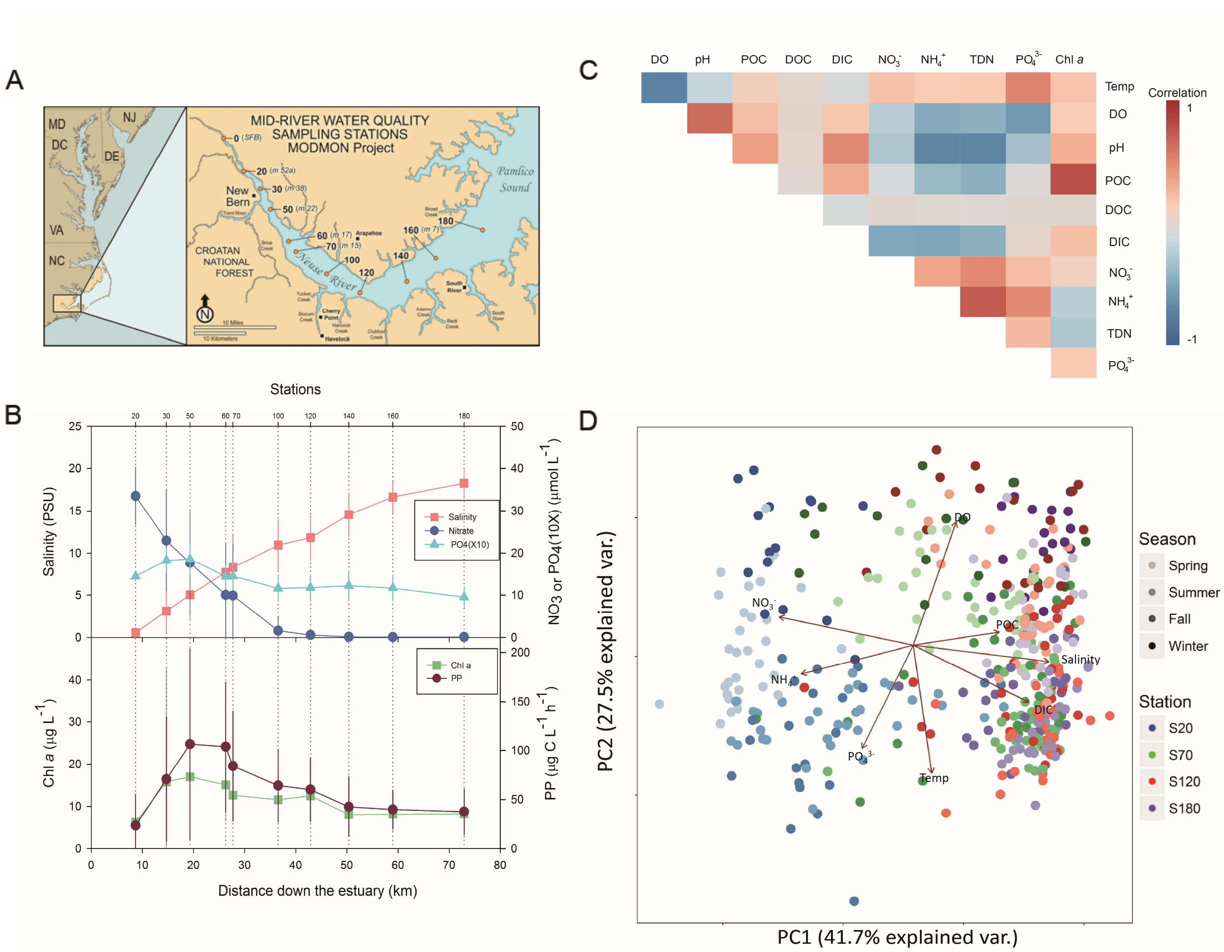


Fig. 1. Physical, chemical and biological properties in the NRE. (A) Location of ModMon stations (20, 70, 120 and 180) for metatranscriptomics sampling in the NRE. (B) Water quality parameters measured in 2012. Plotted are the annual averages and associated standard deviations of surface samples. (C) Correlations between environmental parameters measured in 2012. Temp, surface temperature; DO, dissolved oxygen; POC, particulate organic carbon; DOC, dissolved organic carbon; DIC, dissolved inorganic carbon; NO_3^-, nitrate; NH_4^+, ammonium; TDN, total dissolved nitrogen; PO_4^{3-}, phosphate; Chl *a*, chlorophyll *a*. (D) PCA on environmental measurements from 2007-2012, with stations coded by color and seasons coded by level of shading.

Table 1. Environmental parameters measured during the metatranscriptomic sampling events.

Sampling Time	Station	Temperature (°C)	Salinity	Chl *a* (μg L^{-1})	NO_3^- (μmol L^{-1})	PO_4^{3-} (μmol^{-1})	PPR (mg C $L^{-1}h^{-1}$)
Feb	20	7.74	0.47	6.90	40.07	0.95	5.06
	70	8.64	12.23	26.00	0.05	0.42	100.60
	120	7.85	13.73	9.60	0.03	0.24	20.37
	180	8.44	19.43	4.10	0.00	0.10	10.26
Apr	20	19.46	0.10	5.00	46.64	1.28	18.95
	70	16.68	9.30	10.50	0.00	0.30	75.57
	120	16.98	11.39	9.60	0.00	0.28	61.50
	180	17.48	18.47	7.40	0.00	0.13	54.70
Jun	20	24.31	1.09	6.00	33.93	1.11	50.88
	70	25.09	9.34	12.80	0.00	0.52	52.06
	120	24.09	9.74	12.80	0.03	0.62	71.51
	180	23.53	15.94	6.30	0.05	0.41	41.37
Aug	20	26.67	0.60	7.00	24.79	1.86	22.81
	70	27.17	7.60	16.10	11.29	3.68	220.12
	120	27.03	14.01	10.40	0.00	3.48	130.30
	180	27.54	21.51	5.00	0.00	2.65	56.82
Dec	20	14.76	1.29	2.50	21.29	0.93	3.15
	70	14.68	14.28	12.10	0.00	0.24	48.28
	120	14.42	15.67	6.40	0.00	0.15	28.09
	180	13.21	18.54	3.70	0.00	0.15	13.91

Similarity (ANOSIM) was used to detect spatial/temporal differences in taxonomic composition inferred through transcript abundance of taxonomically assigned contigs in the NRE (Anderson & Walsh 2013). Taxonomic composition was significantly different spatially (p = 0.002) while there was not a significant difference temporally in our 2012 samples (p = 0.405) (Fig. 2B). The similarity percentages program (SIMPER) was then used to identify phytoplankton groups that contribute most to the spatial differences in taxonomic composition (Clarke & Warwick 2001). Diatoms, dinoflagellates, chlorophytes and cryptophytes were found to be the top four groups that make the most contribution to such spatial variation, with transcripts associated with diatoms and dinoflagellates more abundant in the lower estuary and chlorophytes and cryptophytes more abundant in the upper estuary (Fig. 2C).

To detect whether different phytoplankton groups deploy similar strategies to adjust metabolic physiology along the NRE, we applied differential expression analysis to the four dominant eukaryotic phytoplankton functional groups. Of the four groups, gene expression profiles for diatoms were most closely clustered, while chlorophytes and cryptophytes showed the greatest variance (Fig. 2D), which suggests diatoms exhibit a less variable response to changing environmental conditions throughout the estuary when compared to the chlorophytes and cryptophytes.

For most metabolic functions, the four phytoplankton groups shared similar expression patterns along the sampled transect of the NRE. For example, all four groups highly expressed genes involved in central carbohydrate metabolism and carbon fixation at upper estuary sites while genes involved in nitrogen metabolism (NO_3^- assimilation, NO_3^- reduction) were highly expressed at the lower estuary stations. However, there were some metabolic functions that showed distinct patterns among groups (Fig. 2E). For example, chlorophytes, unlike the other main phytoplankton groups, showed increased expression of genes involved in photosynthesis in the lower estuary stations.

Gene expression differences between the upper and lower estuary.

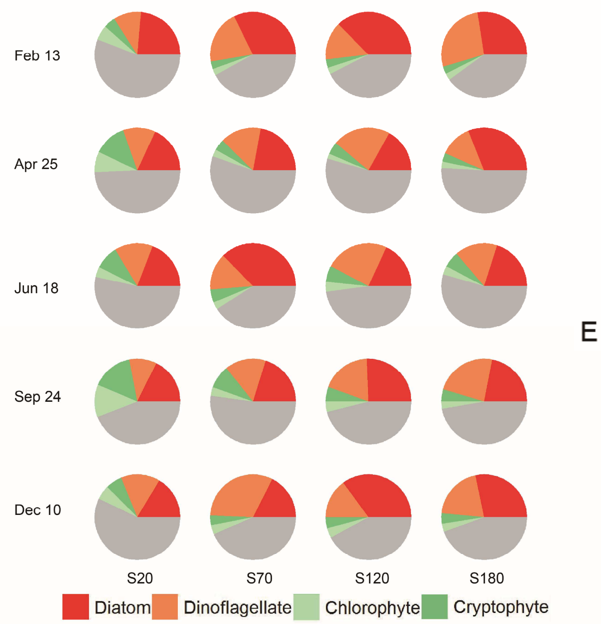
Phytoplankton metabolic physiology can be inferred through gene expression analysis. A PCA was used to examine the differences in gene expression profiles among the different sites and sampling time points. Consistent with environmental factors separating upper and lower sections of the NRE, phytoplankton gene expression in the upper estuary station (St. 20) also differed from the lower estuary stations (Fig. 3A).

Given the high degree of variability found along the NRE stations (Fig .1), we applied a TMM-based differential expression analysis to understand differences in phytoplankton's metabolic physiology between upper and lower estuary stations as perceived through gene expression patterns. Genes with a KEGG Orthology (KO) annotation were grouped into KEGG modules to provide a more holistic analysis of patterns in gene expression. Normalized counts, fold change, and *p*-values for each module are provided in Appendix 3. Photosynthesis (M00161 Photosystem II, M00163 Photosystem I), ATP synthesis (M00158 F-type ATPase, M00160 V-type ATPase, M00162 Cytochrome b6f complex), carbohydrate metabolism (M00001 Glycolysis, M00004 Pentose phosphate pathway, M00009 Citrate cycle), carbon fixation (M00165) and ribosomes (M00177) were over-represented (i.e., increased transcript abundance) at upper stations while at the lower stations, modules for nitrate assimilation (M00615), assimilatory nitrate reduction (M00531), urea transport system (M00323), iron transport system (M00190), N-glycan biosynthesis (M00075), putative multiple sugar transporter (M00207) and putative single sugar transporter (M00221) were over-represented (Fig. 3B). Noticeably, transcripts for Crassulacean

Table 2. Statistics of sequencing, assembly, and quality metrics.

			Assembly		*MarineRef*		*KEGG*	
Sample	**Number Raw Reads**	**N50**	**mapped reads**	**mapped reads%**	**Reads with hits**	**Reads%**	**Reads with functional annotations**	**Reads with functional annotations%**
Sample_1_020	76,546,982	296	47,245,829	61.72%	36,945,820	48.27%	7,641,989	9.98%
Sample_1_070	68,114,590	339	48,679,661	71.47%	40,930,325	60.09%	5,652,406	8.30%
Sample_1_120	98,163,280	414	73,360,652	74.73%	61,556,898	62.71%	9,075,583	9.25%
Sample_1_180	84,547,412	384	48,833,258	57.76%	41,330,582	48.88%	7,541,212	8.92%
Sample_2_020	80,153,398	198	53,408,959	66.63%	40,696,191	50.77%	5,331,874	6.65%
Sample_2_070	77,195,916	355	62,285,449	80.68%	52,091,800	67.48%	8,386,073	10.86%
Sample_2_120	108,465,152	353	87,573,330	80.74%	64,068,014	59.07%	9,469,846	8.73%
Sample_2_180	53,484,164	305	37,044,804	69.26%	2,396,444	4.48%	5,166,408	9.66%
Sample_3_020	67,497,776	229	46,553,365	68.97%	39,333,296	58.27%	4,664,818	6.91%
Sample_3_070	88,682,934	351	64,160,437	72.35%	55,802,712	62.92%	8,042,761	9.07%
Sample_3_120	83,132,158	299	56,390,689	67.83%	44,495,449	53.52%	6,974,708	8.39%
Sample_3_180	62,073,318	213	35,567,596	57.30%	31,625,608	50.95%	4,121,288	6.64%
Sample_4_020	101,367,318	310	80,797,029	79.71%	61,911,608	61.08%	9,183,073	9.06%
Sample_4_070	85,816,112	246	61,162,553	71.27%	46,484,770	54.17%	5,233,581	6.10%
Sample_4_120	123,064,946	450	85,646,123	69.59%	69,918,493	56.81%	10,499,186	8.53%
Sample_4_180	78,494,978	269	46,593,559	59.36%	37,111,577	47.28%	6,263,143	7.98%
Sample_6_020	65,853,028	466	39,711,379	60.30%	31,063,644	47.17%	5,857,446	8.89%
Sample_6_070	98,742,678	476	74,270,668	75.22%	62,047,209	62.84%	8,674,277	8.78%
Sample_6_120	128,778,094	376	128,778,094	79.79%	78,068,697	60.62%	8,982,001	6.97%
Sample_6_180	63,402,458	460	44,421,790	70.06%	34,340,809	54.16%	4,944,770	7.80%

A



B

	Spatial		Temporal	
	R	P	R	P
Taxonomic	0.3317	0.002	0.018	0.405

C

Class	Higher abundance		Contribution
	Upper	Lower	
Diatoms		√	0.1668
Dinoflagellates		√	0.1554
Chlorophytes	√		0.1117
Cryptophytes	√		0.0833

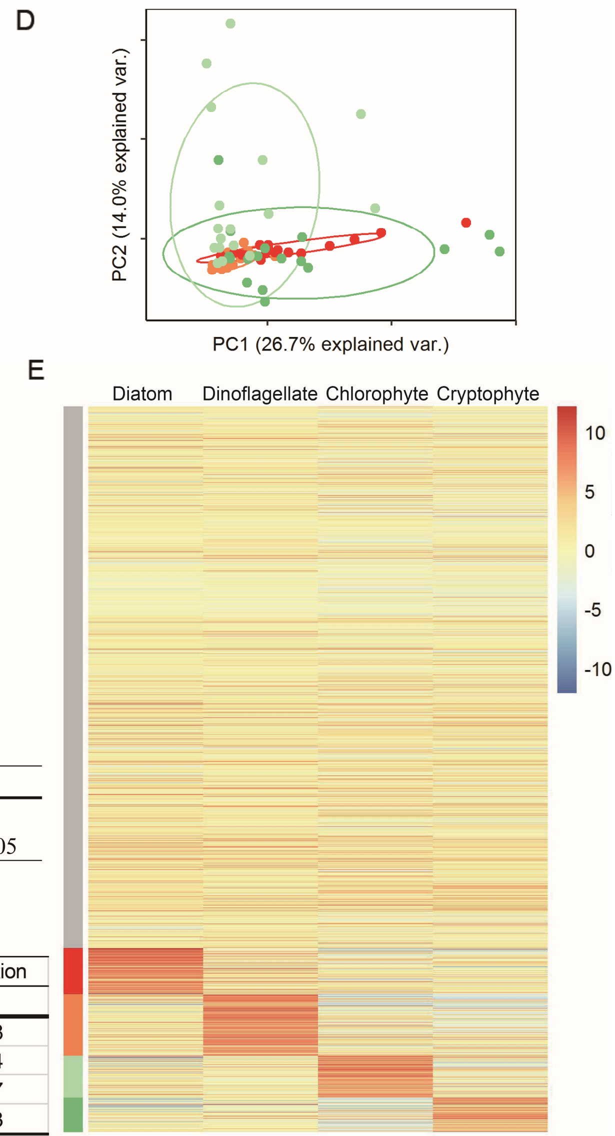


Fig. 2. Taxonomic annotation and patterns of gene expression in the NRE (A) Taxonomic proportions of read counts. Grey indicates all other groups (B) Anosim on taxonomy composition. (C) Simper analysis on taxonomy composition between upper and lower estuary. (D) PCA on gene expression profiles of different phytoplankton, with taxonomy coded by color as shown in A. (E) Heatmap for differential expression analysis between upper and lower estuary stations. Each row represents the expression level of a KO with warm colors (positive fold change) indicating over-representation in the lower estuary stations and cooler colors (negative fold change) indicating under-representation in the lower estuary. KOs were clustered by expression similarity as follows: grey, low variances among differential expression patterns of all four major phytoplankton groups; red, orange, light green, and dark green, KOs that are differentially expressed for diatoms, dinoflagellates, chlorophytes, and cryptophytes, respectively.

acid metabolism (CAM)/C_4 were detected at all sites (Appendix 3). Phosphoenolpyruvate carboxylase (PEPC), which catalyzes the first reaction in CAM/C_4 process,
was over-represented at stations 70 and 120, and underrepresented (i.e., decreased transcript abundance) at station 20.

Correlations between metabolic functions and environmental factors

To assess the correlations between metabolic processes, environmental factors, and sampling sites and time, expression of genes with KO annotations were subjected to Weighted Gene Correlation Network Analysis (WGCNA) (Langfelder & Horvath 2008). In brief, WGCNA clusters KOs into different modules (ME, designated by colors, hereafter denoted subnetworks to avoid confusion with KEGG modules) based on expression dissimilarity, and correlations between each subnetwork and environmental factors were then used to infer relationships between functional KOs and environmental factors (Fig. 4A-B). MEmagenta and MEgreenyellow were positively correlated with nutrient concentrations (surface/bottom NH_4^+, NO_3^- and PO_4^{3-}), and KOs within MEmagenta and MEgreenyellow were highly expressed at station 20 (S20). There are 275 KOs assigned into the MEmagenta group and 352 KOs assigned to MEgreenyellow group. MEred, MEturquiose and MEblack are associated with lower estuary stations 120 and 180 (S120 and S180), and were negatively correlated with nutrient concentrations (Fig. 4A-B). There are 344, 184 and 274 KOs assigned into MEturquiose, MEred and MEbrown, respectively. Genes in each subnetwork encompass a wide spectrum of metabolic functions. A fisher-exact test enrichment analysis was applied to assess which metabolic processes are most representative of each subnetwork. KOs involved in fatty acid metabolism, carbon fixation, photosynthesis, sugar metabolism, ATP synthesis, branched-chain amino acid

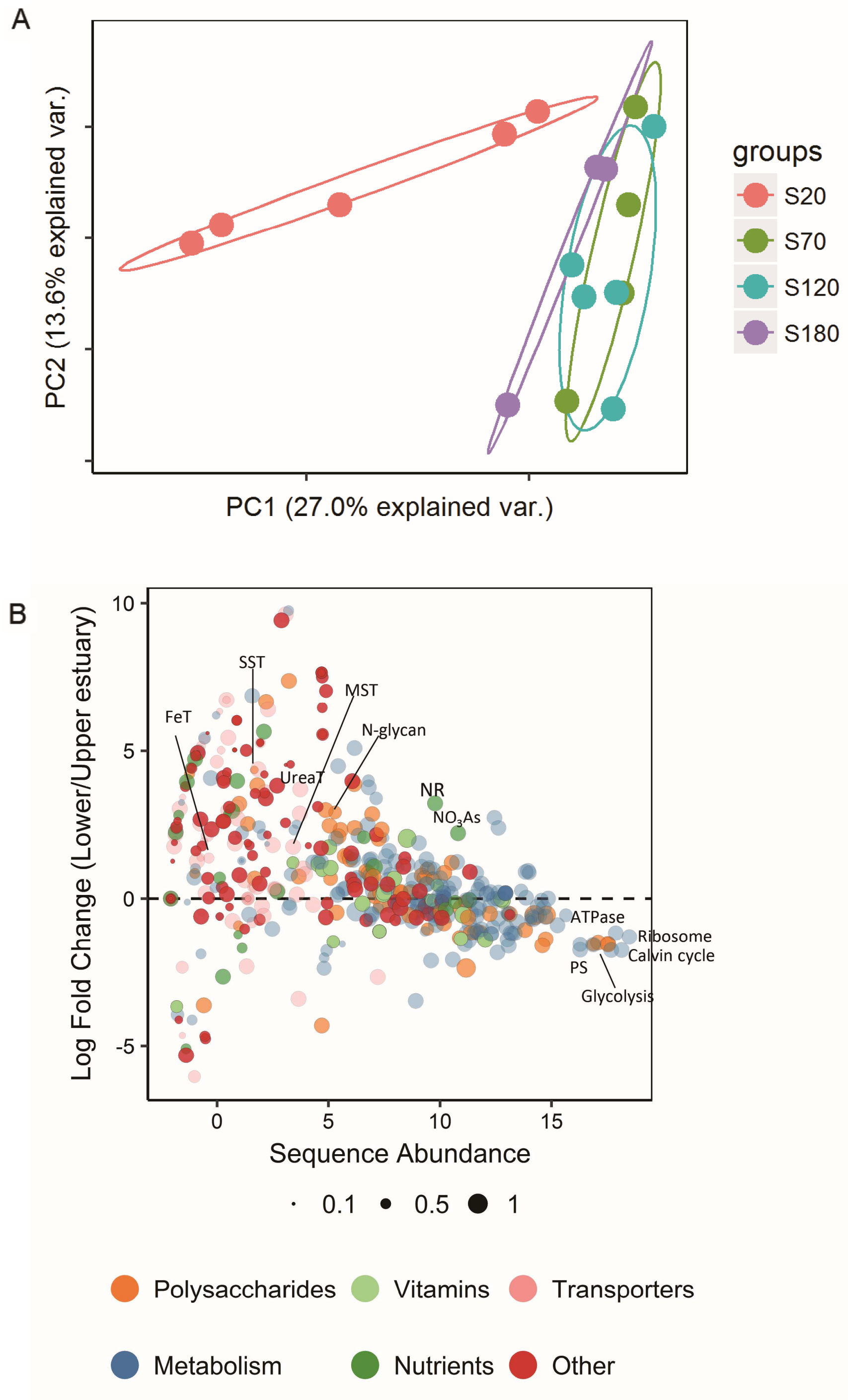


Fig. 3. Gene expression differences between the upper and lower (A) PCA of gene expression profiles of the phytoplankton community grouped by station (B) Metabolic-pathway color-coded scatter plot of log fold-change in transcript abundance between the lower and upper estuary stations. Plotted are the fold-change of module transcripts over average sequence abundance (counts per million [CPM]). Modules are grouped into different categories based on their metabolic functions. Circle size indicates the percentage of enzymes/proteins in each KEGG module with reads mapping to the underlying KEGG genes. Both axes are in log scale (base 2).

metabolism and ribosome metabolism are enriched in MEmagenta and MEgreenyellow (Fig. 4C and Table 3). KOs representing nitrogen metabolism, lipid metabolism, glycan metabolism, transcription/translation regulation mechanisms (spliceosome, RNA processing, proteasome), ABC transport system and saccharide, polyol and lipid transport system are enriched in MEred, MEturquiose and MEblack (Fig. 4C and Table 3).

In accordance with the KOs that are represented within each of these subnetworks (i.e., MEmagenta, MEgreenyellow, MEred, MEturquiose and MEblack), the phytoplankton responses to variations in environmental gradients along the NRE can be primarily characterized as being related to nutrient metabolism, growth metabolism, saccharide metabolism, and transport systems, grouping greater than 9500 KOs into 20 modules. Pearson correlation analysis was applied to determine how those "responsive" modules correlate with environmental factors. A negative correlation was detected between nitrogen concentrations and modules enriched with genes involved in nitrogen-related processes (i.e., NO_3^- reduction, NO_3^- assimilation, NO_3^- transporter, urea transport system), ABC transport system and transporters for essential metabolites such as iron, multiple sugars, proline, phosphate and glycose (Fig. 5A). Nutrient concentrations (NO_3, surface/bottom NH_4, and PO_4) were positively correlated with modules enriched with genes involved in photosynthesis (PSI, PSII), glycolysis, photorespiration, ribosome metabolism, fatty acid synthesis, and citrate cycle (Fig. 5A). Besides correlation between modules and environmental factors, correlations within modules were also examined. Positive correlations among modules representative of transporters for various nutrients, modules involved in nitrogen-related processes and lipopolysaccharide synthesis/releasing system were identified (Fig. 5B). Photosynthesis, carbohydrate metabolisms and ATP synthesis were also detected to be co-expressed (Fig. 5C).

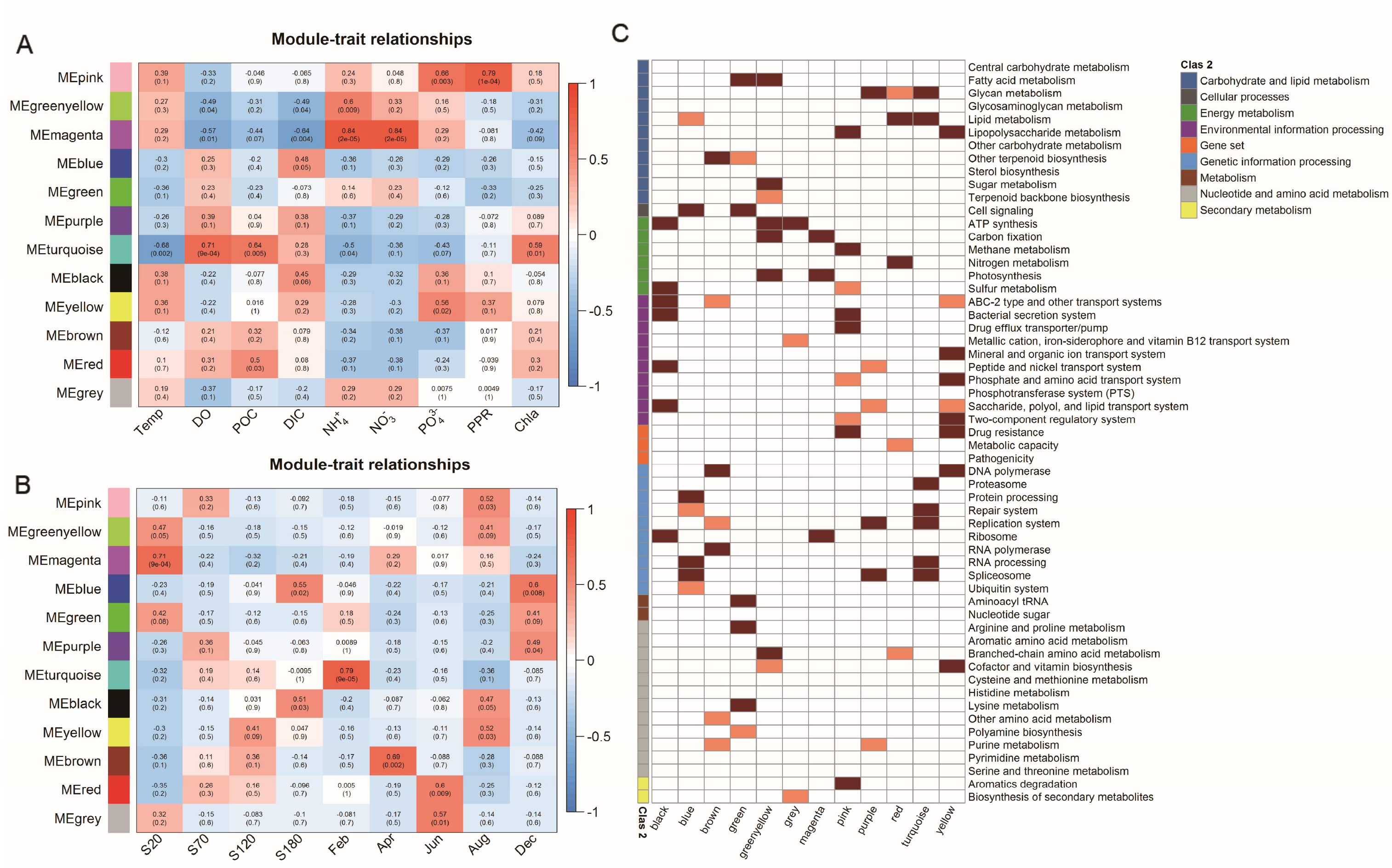


Fig. 4. WGCNA and gene enrichment analysis. Weighted Gene Co-expression Network Analysis (WGCNA) demonstrating the correlation between subnetworks and (A) environmental measurements (B) spatiotemporal identifiers. The numbers in each cell are Pearson's correlation coefficients and *p*-values of the correlation test (in brackets). Color of the cell indicates the correlation between subnetworks and corresponding parameters, with red indicating a positive correlation and blue indicating a negative correlation. (C) Heatmap for enrichment analysis shows the significance levels of whether each KEGG Clas3 is enriched in each subnetwork, with dark red indicative of p-value < 0.05, and pink indicative of 0.05<p-value < 0.1. KEGG Clas3 functions are sorted by KEGG Clas2 in alphabetic order.

Table 3: Enrichment analysis on KEGG Clas3

KEGG Clas3	P-value	subnetwork
Sugar metabolism	0.031498	greenyellow
Carbon fixation	0.025377	greenyellow
Photosynthesis	0.032296	greenyellow
ATP synthesis	5.16E-07	greenyellow
Fatty acid metabolism	0.041549	greenyellow
Branched-chain amino acid metabolism	0.000373	greenyellow
Ribosome	2.25E-51	magenta
Carbon fixation	0.023032	magenta
Photosynthesis	0.00084	magenta
RNA processing	0.000458	turquoise
Spliceosome	0.001091	turquoise
Proteasome	1.32E-21	turquoise
Repair system	1.18E-06	turquoise
Replication system	0.005877	turquoise
Lipid metabolism	0.025833	turquoise
Glycan metabolism	0.049328	turquoise
Bacterial secretion system	0.001129	black
ABC-2 type and other transport systems	0.017615	black
Peptide and nickel transport system	1.41E-05	black
Saccharide, polyol, and lipid transport system	5.20E-06	black
Ribosome	0.029372	black
Sulfur metabolism	0.000451	black
ATP synthesis	0.001319	black
Cell signaling	7.00E-05	green
Aminoacyl tRNA	3.66E-05	green
Fatty acid metabolism	0.025243	green
Lysine metabolism	0.009661	green
Arginine and proline metabolism	0.021213	green
DNA polymerase	0.001355	brown
RNA polymerase	0.032672	brown
Other terpenoid biosynthesis	0.028148	brown
Drug efflux transporter/pump	0.00282	pink
Drug resistance	0.006899	pink
Aromatics degradation	0.017968	pink
Bacterial secretion system	0.007337	pink
Methane metabolism	0.00085	pink
Lipopolysaccharide metabolism	0.01699	pink
Cell signaling	0.000145	blue
Protein processing	0.043808	blue
RNA processing	6.02E-09	blue

Spliceosome	1.06E-07	blue
Nitrogen metabolism	0.004075	red
Lipid metabolism	0.023006	red
ATP synthesis	0.000208	grey
Drug resistance	0.004657	yellow
Two-component regulatory system	0.000763	yellow
DNA polymerase	0.015433	yellow
Phosphate and amino acid transport system	0.000124	yellow
Mineral and organic ion transport system	0.004157	yellow
Cofactor and vitamin biosynthesis	0.017866	yellow
Lipopolysaccharide metabolism	0.006755	yellow
Spliceosome	0.005551	purple
Replication system	0.003774	purple
Glycan metabolism	0.011045	purple

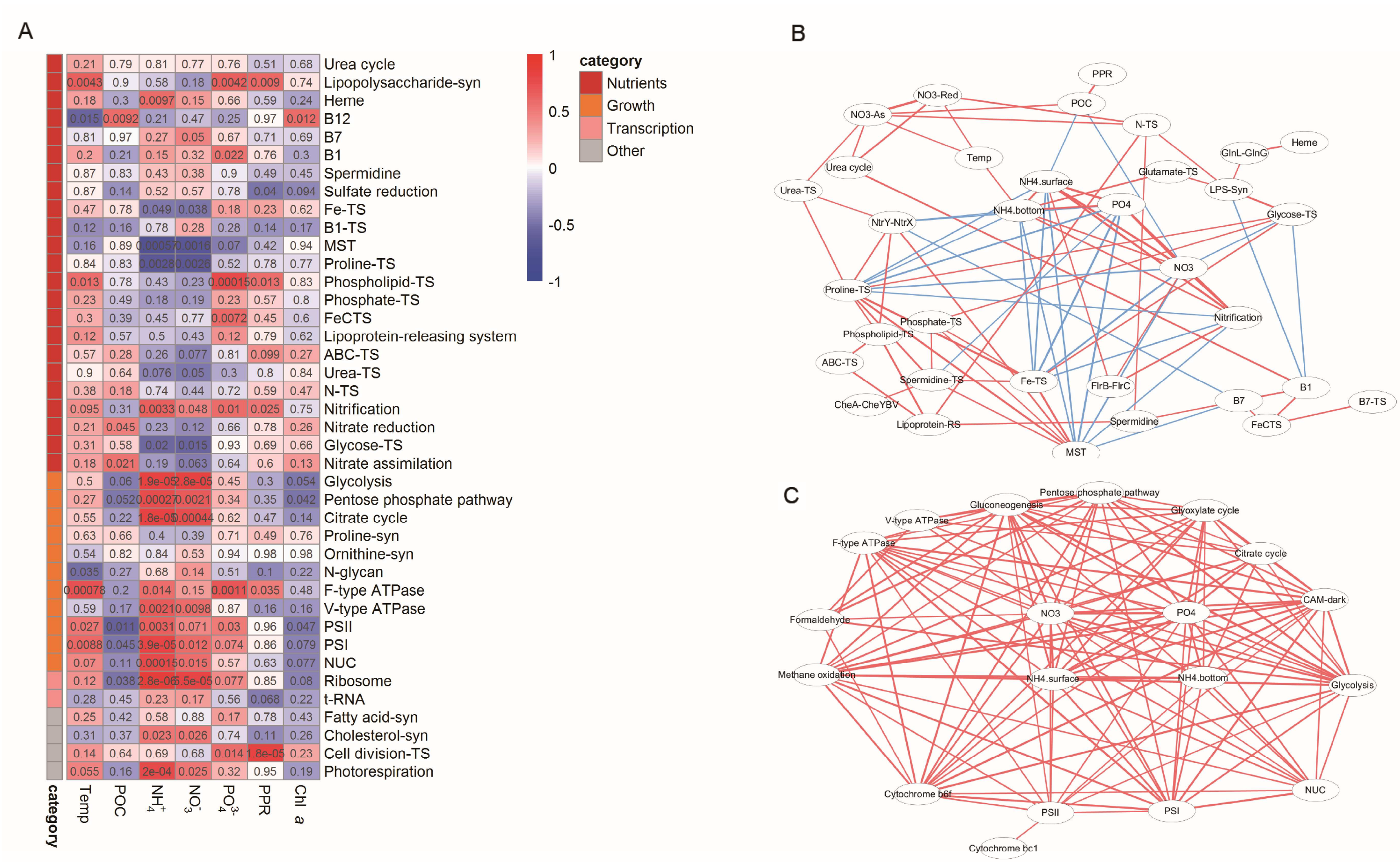


Fig. 5. Correlation and network analysis (A) Heatmap of the correlation between KEGG Modules and selected environmental factors in the NRE. Color of the cell indicates the correlation where red indicates a positive correlation and blue indicates a negative correlation. The numbers in the cells are p-values of the Pearson correlation test. Modules are sorted based on functional categories. Selected network map for (B) nutrient metabolisms and (C) growth-related metabolisms showing the co-expression patterns in KEGG modules and correlations between modules and environmental factors. Connections stand for significant (p-value < 0.05) correlations, with red and blue indicative of positive and negative correlations, respectively.

DISCUSSION

The NRE is a highly dynamic system characterized by steep gradients in NO_3^- concentrations from the upper to lower estuary (Fig. 1B). Correspondingly, nitrogen has frequently been shown to be the major limiting resource for phytoplankton with NO_3^- being the dominant inorganic source in the NRE (Piehler et al. 2004; Cira et al. 2016). Nitrogen is essential for phytoplankton growth, but nitrogen inputs are dependent on discharge rate and can be quite episodic. For example, the annual average of nitrate concentration at station 70 is approximately 10 μmol L^{-1}, although concentrations were near or below the level of detection during four out of five metatranscriptiome sampling time points (Table 1). Given the high variability of nitrogen inputs and strong interplay between biological and hydrological processes, nitrogen (NO_3^- and NH_4^+) along with other environmental factors were only weakly correlated with chl *a* (Fig. 1C), and thus models based solely on these bulk measurements would have difficulty in predicting phytoplankton dynamics within the NRE.

Environmental conditions were found to be relatively similar at the three lower estuary stations (station 70, 120 and 180) while the upper estuary station (station 20) often displayed the greatest differences in water properties (Fig. 1D). Concomitantly, metabolic activity inferred through gene expression patterns also displayed many differences between the upper and lower estuary stations (e.g. Fig. 3A).

Transcript proportions of chlorophytes and cryptophytes decrease from upper to lower estuary, whereas those of dinoflagellates and diatoms slightly increase along the estuary. This spatial segregation of taxonomic groups is consistent with previous findings that freshwater chlorophytes are likely to be more abundant in the upper estuary due to salinity preferences (Pinckney et al. 1997; Paerl et al. 1995; Valdes-Weaver et al. 2006). Percentages of diatom transcripts in the NRE are relatively stable. Consistent with this pattern, there were minimal variations in diatom gene expression along the NRE despite the steep spatial gradients observed (Fig. 2D). Diatoms are known to have high maximum intrinsic growth rates and can become the dominant group when nutrients are supplied (Smayda 1997). Similarly within coastal upwelling zones as well as during nutrient addition experiments, diatoms often elicit a substantial gene expression response and quickly dominate the stimulated phytoplankton community (Fawcett & Ward n.d.; Smith et al. 1992; Marchetti et al. 2012). However, in the NRE, diatoms were not any more abundant than other phytoplankton groups at upper estuary where nutrient concentrations were high, and diatom blooms were not observed frequently despite the occurrence of dinoflagellate blooms in the NRE when there are high nutrient inputs (Pinckney *et al.*, 1998; Gong *et al*. 2016). Hall *et al.*, (2012) reported resuspended diatoms from benthic environments as a major component to the diatom community that only weakly relied on “new” riverine nutrient inputs. Thus the diatom community could be inherently less affected by the spatial gradients in nutrient concentrations resulting in more stable gene expression patterns.

Divergence in gene expression profiles between upper and lower estuary stations indicated that phytoplankton in these sectors displayed different metabolic activities. Photosynthesis, ATP synthesis, carbohydrate metabolism, and carbon fixation were over-represented in the upper estuary station, all of which are important components for phytoplankton energy production and growth. Ribosomes, which are essential for protein synthesis, were also over-represented at station 20. The amount of ribosome transcripts has previously been shown to be positively correlated with metabolic rates and growth (Gifford et al.

2012; Wei et al. 2001). The over-representation of these modules suggests high phytoplankton growth rates at this station, which is consistent with previous findings based on nutrient uptake kinetics that the highest growth rates occur in this section of estuary despite phytoplankton biomass not typically accumulating due to significant advective losses induced by high discharge rates (Pinckney et al. 1997). At lower estuary stations, nitrogen-related modules were over-represented, indicative of a high nitrogen demand (Fig. 3B).

CAM and C_4 are CO_2 concentrating mechanisms (CCM) acquired by algae for carbon fixation to cope with high O_2 concentrations relative to CO_2 (Giordano et al. 2005; Reinfelder et al. 2000). Phosphoenolpyruvate carboxylase (PEPC) catalyzes the first reaction in CAM/C_4 process, incorporating inorganic carbon to phosphoenolpyruvate (PEP) to form the four-carbon compound oxaloacetate, which is used to release CO_2 for carbon fixation in downstream reactions (Hatch & Slack 1968). Transcripts for PEPC were over-represented at stations 70 and 120, and were under-represented at station 20. As phytoplankton consume CO_2 and produce O_2, the high biomass at stations 70 and 120 further increases the O_2/CO_2 ratio, which could explain the high expression of PEPC. Malate dehydrogenase (oxaloacetate-decarboxylating) catalyzes the release of CO_2 so that CO_2 can be utilized in the Calvin cycle. Malate dehydrogenases (oxaloacetate-decarboxylating) (K00025, K00026, K00029) were over-represented at station 20, indicative of high CO_2 assimilation in the upper estuary phytoplankton community, consistent with high carbon fixation and high growth rates.

Energy production, carbon fixation and other growth metabolisms that were highly expressed at upper estuary were represented by WGCNA subnetworks MEmagenta and MEgreenyellow (Fig. 4C, Table. 3). These subnetworks had strong positive correlations with nitrogen concentrations (NO_3^-, NH_4^+), thus growth metabolisms in general were positively related to nitrogen concentrations. Correlations between each individual module involved in growth metabolisms (including PSI, PSII, carbohydrate metabolism, ATP synthesis and ribosome metabolisms) and nitrogen concentrations were found to be strongly positive (Fig. 5A). Such positive correlations indicate nitrogen is essential for phytoplankton growth and suggests the high nitrogen concentrations explain why growth-related modules were highly expressed in the upper estuary.

In contrast, nitrogen-related processes (i.e., NO_3^- assimilation, urea transport system, NO_3^- reduction) were enriched in subnetwork MEred (Fig. 4C) and found to be negatively correlated with nitrogen concentration (i.e., NO_3^-, NH_4^+) (Fig. 5A). The negative correlation between these subnetworks and nitrogen concentrations is consistent with previous findings that have shown genes for nitrogen metabolism increase in expression when nitrogen is limiting (Maheswari et al. 2010; Kang et al. 2007; Song & Ward 2007). Similarly, transporters for iron, multiple sugars, proline, phosphate, glycose were also negatively correlated with nitrogen concentrations, indicating there were high demands for those essential metabolites in the lower estuary. Co-expression patterns were detected among nitrogen utilization pathways, which indicates nitrogen-related metabolic functions were coordinated under nitrogen stress conditions. Other than nitrogen metabolism, transcription/translation activities (replication system, repair system, DNA polymerase) and ABC transport systems were also enriched in subnetworks (MEturquiose, MEblack) that were negatively correlated with nitrogen concentrations. Enrichment of transcription/translation activities can be another signal that phytoplankton were under nutrient stress as these processes have been shown to function as recycling mechanisms for protein and amino acids under nutrient stress (Liu *et al.*, 2015). The ABC transport system is comprised of transmembrane transporters and are essential for saccharides export (Jones and

George, 2004; Delbarre-Ladrat *et al.*, 2014). Phytoplankton have been shown to export carbohydrate for multiple reasons, including balancing inner C/N ratios and in exchange for remineralized nutrients from bacteria (Wear *et al.*, 2015; Croft *et al.*, 2005; Rinta-Kanto *et al.*, 2012; Cooper and Smith, 2015; Azam *et al.*, 1983; Gong *et al.*, 2016). As illustrated above, phytoplankton have high demands for N, P and Fe in the lower estuary, thus elevated export of carbohydrate can be a key process to sustain surrounding bacteria growth for remineralized ammonium and enhancement of Fe availability (Seymour *et al.*, 2017). Thus, enrichment of glycan metabolism, lipopolysaccharide metabolism and ABC transport system and the strong negative correlation between sugar transportation (multiple sugar transporter, ABC transport system) could infer there are mutualistic nutrients/carbon exchange mechanisms between phytoplankton and bacteria in the lower estuary. In addition, considering a limited N supply along with replete carbon sources, it is also reasonable for phytoplankton to export excess carbon-rich substrates to balance C/N ratio.

Summary

Through combining gene expression profiling with measurements of environmental conditions and other water quality parameters, we have constructed a more comprehensive network analysis for the plankton community in the NRE. Phytoplankton implement different metabolic strategies in the upper and lower estuary as a consequence of the steep environmental gradients that persist in the NRE. In the upper estuary, where nutrients are replete, phytoplankton highly expressed photosynthesis, carbon fixation, and other growth-related metabolic pathways. While in the lower estuary, where nutrients are limiting, phytoplankton increase nutrient acquisition processes to scavenge the scarce nutrients from the environment and elevate their internal nutrient cycling to meet the high demands for various nutrients. Metabolisms for polysaccharide synthesis and transportation were elevated in lower estuary and could be reflective of unbalanced growth and/or interactions with their surrounding microbial consortia. Nitrogen concentrations are shown to be the main drivers for the divergence in the phytoplankton differences in gene expression between the upper and lower estuaries. Although NO_3^- concentrations were found to be mostly limiting at Station 70, occasionally, high discharge events can deliver large amount of NO_3^- to station 70 resulting in blooms that may alter gene expression patterns to be more reflective of the upper estuary phytoplankton community. Sampling of such instances would enable a better understanding of the phytoplankton dynamics in the NRE.

The sensitivity of phytoplankton metabolic functional responses to their environment as elucidated through molecular approaches has provided us with a new tool for water quality monitoring within our valuable coastal regions. With the development of sequencing technology and corresponding bioinformatics tools, more informative models can be developed to better predict the consequences of changes in environmental factors on phytoplankton dynamics and the marine ecosystem.

APPENDIX 1: SRA ACCESSION NUMBERS FOR METATRANSCRIPTOMIC SAMPLES

Sample	Study accession no.
Sample_1_020	SRR3218234
Sample_1_070	SRR3218378
Sample_1_120	SRR3219836
Sample_1_180	SRR3221672
Sample_2_020	SRR3233860
Sample_2_070	SRR3233861
Sample_2_120	SRR3233862
Sample_2_180	SRR3233863
Sample_3_020	SRR3233866
Sample_3_070	SRR3233867
Sample_3_120	SRR3233868
Sample_3_180	SRR3233869
Sample_4_020	SRR3233870
Sample_4_070	SRR3233886
Sample_4_120	SRR3233887
Sample_4_180	SRR3233926
Sample_6_020	SRR3234059
Sample_6_070	SRR3234060
Sample_6_120	SRR3234061
Sample_6_180	SRR3234482

APPENDIX 2: CUSTOM TAXONOMIC LOOK-UP TABLE

Class	Custom Name
Coscinodiscophyceae	Diatoms
Chrysophyceae	Chrysophyte
Dinophyceae	Dinoflagellates
Spirotrichea	Ciliates
Synurophyceae	Synurophytes
Cryptophyta	Cryptophytes
Prymnesiaceae	Haptophytes
Fragilariophyceae	Diatoms
Mamiellophyceae	Prasinophytes
Chlorophyta	Chlorophycean
Mediophyceae	Diatoms
Litostomatea	Ciliates
Dictyochophyceae	Dictyochophytes
Chlorophyceae	Chlorophycean
Foraminifera	Foramnifera
Bacillariophyceae	Diatoms
Oligohymenophorea	Ciliates
Heterotrichea	Ciliates
Trebouxiophyceae	Trebouxiophytes
Phaeocystaceae	Haptophytes
Raphidophyceae	Raphidophytes
Pelagophyceae	Pelagophytes
Prostomatea	Ciliates
Thraustochytriaceae	Labyrinthulids
Coccolithaceae	Haptophytes
Colpodea	Ciliates
Glaucocystophyceae	Glaucophytes
Noelaerhabdaceae	Haptophytes
Chromerida	Chromera
Chlorodendrophyceae	Chlorophycean
Codonosigidae	Choanoflagellates
Pavlovaceae	Haptophytes
Alphaproteobacteria	
Acanthoecidae	Choanoflagellates
Isochrysidaceae	Haptophytes
Bangiophyceae	Bangiophytes
Xanthophyceae	Xanthophytes
Synchromophyceae	
Flavobacteriia	
Bolidophyceae	Bolidophytes

Agaricomycetes	Basidiomycetes
Gammaproteobacteria	
Chrysochromulinaceae	Haptophytes
Compsopogonophyceae	
Paramoebidae	Dactyiopodids
Verrucomicrobia	
Cyanobacteria	
Cafeteriaceae	Bicosoecids
Nephroselmidophyceae	Prasinophytes
Euglenida	Euglenids
Pleurochrysidaceae	Haptophytes
Vannellidae	Vannellids
Betaproteobacteria	
Echinamoebidae	Tubulinids
Streptophyta	
Calcidiscaceae	Haptophytes
Eustigmatophyceae	Eustigmatophytes
Heterolobosea	Heteroloboseans
Pinguiophyceae	Pinguiophytes
Rhodellophyceae	
Stylonematophyceae	
Planctomycetia	
Vexilliferidae	Dactyiopodids
Euryarchaeota	
Sphingobacteriia	
Perkinsidae	Perkinsus
Saccharomycetes	Ascomycetes
Thaumarchaeota	
Actinobacteria	
Cytophagia	
Difflugiidae	Tubulinids
Bodonidae	Euglenids
Gregarinasina	Apicomplexa
Haplosporidia	Haplosporidia
Lentisphaeria	
Thermoplasmata	
Marinimicrobia	

APPENDIX 3: RESULT OF DIFFERENTIAL EXPRESSION ANALYSIS

module	upper	lower	logFC	logCPM	name	clas3	pvalue
M00001	281726	97625	-1.529	17.5332	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	Central carbohydrate metabolism	3.62E-07
M00002	206705	73424.6	-1.4932	17.0957	Glycolysis, core module involving three-carbon compounds	Central carbohydrate metabolism	1.70E-07
M00003	286768	97031.2	-1.5634	17.55	Gluconeogenesis, oxaloacetate => fructose-6P	Central carbohydrate metabolism	5.33E-12
M00004	39678.3	15250.7	-1.3795	14.7453	Pentose phosphate pathway (Pentose phosphate cycle)	Central carbohydrate metabolism	8.99E-10
M00005	134.274	141.189	0.07244	7.10571	PRPP biosynthesis, ribose 5P => PRPP	Central carbohydrate metabolism	0.90272
M00006	892.197	2056.69	1.20489	10.526	Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P	Central carbohydrate metabolism	0.0128
M00007	37115.7	12332.5	-1.5896	14.5936	Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P	Central carbohydrate metabolism	1.79E-09
M00008	470.645	924.86	0.9746	9.44657	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	Central carbohydrate metabolism	0.32818
M00009	33637.6	22990.8	-0.549	14.7892	Citrate cycle (TCA cycle, Krebs cycle)	Central carbohydrate metabolism	0.05221
M00010	10596.2	7375.16	-0.5228	13.1334	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	Central carbohydrate metabolism	0.1814
M00011	23041.4	15615.6	-0.5612	14.2384	Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate	Central carbohydrate metabolism	0.089
M00012	19558.3	9361.02	-1.063	13.8197	Glyoxylate cycle	Other carbohydrate metabolism	0.00128
M00013	1028.65	898.32	-0.1955	9.91212	Malonate semialdehyde pathway, propanoyl-CoA => acetyl-CoA	Other carbohydrate metabolism	0.99542
M00014	4899.16	2195.42	-1.158	11.7925	Glucuronate pathway (uronate pathway)	Other carbohydrate metabolism	0.02543
M00015	159.569	246.629	0.62816	7.66604	Proline biosynthesis, glutamate => proline	Arginine and proline metabolism	0.10156
M00016	4252.05	2001.6	-1.087	11.6105	Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine	Lysine metabolism	0.08863
M00017	3329.21	1821.4	-0.8701	11.3305	Methionine biosynthesis, apartate => homoserine => methionine	Cysteine and methionine metabolism	0.01929
M00018	3238.12	1449.9	-1.1592	11.1948	Threonine biosynthesis, aspartate => homoserine => threonine	Serine and threonine metabolism	0.02442
M00019	4936.64	2167.52	-1.1875	11.7944	Valine/isoleucine biosynthesis, pyruvate => valine / 2-oxobutanoate => isoleucine	Branched-chain amino acid metabolism	0.25961
M00020	1161.86	1477.15	0.34638	10.3658	Serine biosynthesis, glycerate-3P => serine	Serine and threonine metabolism	0.56789
M00021	409.583	372.567	-0.1367	8.6113	Cysteine biosynthesis, serine => cysteine	Cysteine and methionine metabolism	0.05428
M00022	988.933	680.255	-0.5398	9.70493	Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismate	Aromatic amino acid metabolism	0.22741
M00023	582.168	513.104	-0.1822	9.09707	Tryptophan biosynthesis, chorismate => tryptophan	Aromatic amino acid metabolism	0.48835
M00024	87.3266	64.2596	-0.4425	6.244	Phenylalanine biosynthesis, chorismate => phenylalanine	Aromatic amino acid metabolism	0.96544
M00025	121.807	146.253	0.26387	7.06641	Tyrosine biosynthesis, chorismate => tyrosine	Aromatic amino acid metabolism	0.9663
M00026	267.196	270.587	0.0182	8.07088	Histidine biosynthesis, PRPP => histidine	Histidine metabolism	0.91699
M00027	152.042	362.426	1.25321	8.00694	GABA (gamma-Aminobutyrate) shunt	Other amino acid metabolism	0.62037
M00028	232.134	199.519	-0.2184	7.75373	Ornithine biosynthesis, glutamate => ornithine	Arginine and proline metabolism	0.50052
M00029	755.91	699.246	-0.1124	9.50696	Urea cycle	Arginine and proline metabolism	0.52802

M00030	365.215	239.643	-0.6079	8.24045	Lysine biosynthesis, AAA pathway, 2-oxoglutarate => 2-aminoadipate => lysine	Lysine metabolism	0.99906
M00031	1.31453	0.86361	-0.6061	0.1231	Lysine biosynthesis, mediated by LysW, 2-aminoadipate => lysine	Lysine metabolism	0.77096
M00032	3110.92	2209.58	-0.4936	11.3773	Lysine degradation, lysine => saccharopine => acetoacetyl-CoA	Lysine metabolism	0.79761
M00033	2475.94	588.285	-2.0734	10.5813	Ectoine biosynthesis, aspartate => ectoine	Other amino acid metabolism	3.03E-05
M00034	22831.9	14306.3	-0.6744	14.1806	Methionine salvage pathway	Cysteine and methionine metabolism	0.31296
M00035	26935.1	18198.5	-0.5657	14.4619	Methionine degradation	Cysteine and methionine metabolism	0.18287
M00036	5343.58	4496.34	-0.2491	12.2644	Leucine degradation, leucine => acetoacetate + acetyl-CoA	Branched-chain amino acid metabolism	0.9907
M00037	10.164	8.94916	-0.1836	3.2565	Melatonin biosynthesis, tryptophan => serotonin => melatonin	Aromatic amino acid metabolism	0.92635
M00038	76.7321	168.456	1.13447	6.93775	Tryptophan metabolism, tryptophan => kynurenine => 2-aminomuconate	Aromatic amino acid metabolism	0.8509
M00039	70.0214	289.619	2.04829	7.49041	Monolignol biosynthesis, phenylalanine/tyrosine => monolignol	Biosynthesis of secondary metabolites	0.0537
M00040	48.4494	61.6981	0.34875	5.78329	Tyrosine biosynthesis, prephanate => pretyrosine => tyrosine	Aromatic amino acid metabolism	0.13069
M00042	156.658	278.496	0.83003	7.76538	Catecholamine biosynthesis, tyrosine => dopamine => noradrenaline => adrenaline	Aromatic amino acid metabolism	0.47539
M00044	217.902	256.272	0.234	7.88927	Tyrosine degradation, tyrosine => homogentisate	Aromatic amino acid metabolism	0.74639
M00045	12.0302	27.8771	1.21242	4.31858	Histidine degradation, histidine => N-formiminoglutamate => glutamate	Histidine metabolism	0.39537
M00046	148.663	101.816	-0.5461	6.96854	Pyrimidine degradation, uracil => beta-alanine, thymine => 3-aminoisobutanoate	Pyrimidine metabolism	0.6943
M00047	1096.48	1012.41	-0.1151	10.0423	Creatine pathway	Other amino acid metabolism	0.45958
M00048	659.764	1037.91	0.65366	9.72934	Inosine monophosphate biosynthesis, PRPP + glutamine => IMP	Purine metabolism	0.6699
M00049	14468.1	6408.82	-1.1747	13.3496	Adenine ribonucleotide biosynthesis, IMP => ADP,ATP	Purine metabolism	0.11932
M00050	13405.1	5871.7	-1.1909	13.2346	Guanine ribonucleotide biosynthesis IMP => GDP,GTP	Purine metabolism	0.13419
M00051	234.821	518.962	1.14407	8.55801	Uridine monophosphate biosynthesis, glutamine (+ PRPP) => UMP	Pyrimidine metabolism	0.34649
M00052	9644.88	2713.23	-1.8298	12.5932	Pyrimidine ribonucleotide biosynthesis, UMP => UDP/UTP,CDP/CTP	Pyrimidine metabolism	0.01041
M00053	10794.6	5465.91	-0.9818	12.9891	Pyrimidine deoxyribonuleotide biosynthesis, CDP/CTP => dCDP/dCTP,dTDP/dTTP	Pyrimidine metabolism	0.00126
M00055	94.3028	150.27	0.67219	6.93412	N-glycan precursor biosynthesis	Glycan metabolism	0.29252
M00056	30.8445	84.2602	1.44984	5.8468	O-glycan biosynthesis, mucin type core	Glycan metabolism	0.22431
M00057	9.51106	15.9151	0.74272	3.66824	Glycosaminoglycan biosynthesis, linkage tetrasaccharide	Glycosaminoglycan metabolism	0.62346
M00058	0.46228	6.59144	3.83377	1.81838	Glycosaminoglycan biosynthesis, chondroitin sulfate backbone	Glycosaminoglycan metabolism	0.4496
M00059	48.0671	34.5697	-0.4755	5.36871	Glycosaminoglycan biosynthesis, heparan sulfate backbone	Glycosaminoglycan metabolism	0.93744
M00060	26.4393	42.286	0.6775	5.10277	Lipopolysaccharide biosynthesis, KDO2-lipid A	Lipopolysaccharide metabolism	0.75181
M00061	9.16631	134.244	3.87237	6.164	D-Glucuronate degradation, D-glucuronate => pyruvate + D-glyceraldehyde 3P	Other carbohydrate metabolism	0.16524
M00063	30.2795	219.363	2.85691	6.96372	CMP-KDO biosynthesis	Lipopolysaccharide metabolism	0.01951
M00064	2.50053	3.97471	0.66862	1.69493	ADP-L-glycero-D-manno-heptose biosynthesis	Lipopolysaccharide metabolism	0.61363
M00065	176.936	149.965	-0.2386	7.35271	GPI-anchor biosynthesis, core oligosaccharide	Glycan metabolism	0.21127

M00066	0.3584	0.64491	0.84751	-0.9952	Lactosylceramide biosynthesis	Lipid metabolism	0.98543
M00067	0.11285	18.6554	7.36898	3.23022	Sulfoglycolipids biosynthesis, ceramide/1-alkyl-2-acylglycerol => sulfatide/seminolipid	Lipid metabolism	0.53011
M00068	2.5459	1.34624	-0.9192	0.96057	Glycosphingolipid biosynthesis, globo-series, LacCer => Gb4Cer	Glycan metabolism	0.60911
M00069	0.04356	0.80911	4.21528	-1.2299	Glycosphingolipid biosynthesis, ganglio series, LacCer => GT3	Glycan metabolism	0.81109
M00070	0.39173	3.61043	3.20424	1.00078	Glycosphingolipid biosynthesis, lacto-series, LacCer => Lc4Cer	Glycan metabolism	0.99383
M00071	0.7778	4.46949	2.52264	1.39157	Glycosphingolipid biosynthesis, neolacto-series, LacCer => nLc4Cer	Glycan metabolism	0.9922
M00072	667.012	506.07	-0.3984	9.19609	N-glycosylation by oligosaccharyltransferase	Glycan metabolism	0.09578
M00073	283.296	322.841	0.18851	8.2435	N-glycan precursor trimming	Glycan metabolism	0.92991
M00074	275.677	324.605	0.23571	8.2295	N-glycan biosynthesis, high-mannose type	Glycan metabolism	0.90589
M00075	9.20599	69.515	2.91668	5.29868	N-glycan biosynthesis, complex type	Glycan metabolism	0.39687
M00076	15.727	77.7502	2.3056	5.54654	Dermatan sulfate degradation	Glycosaminoglycan metabolism	0.95419
M00077	10.1579	56.0668	2.46454	5.0493	Chondroitin sulfate degradation	Glycosaminoglycan metabolism	0.87684
M00078	16.6455	71.0124	2.09294	5.45381	Heparan sulfate degradation	Glycosaminoglycan metabolism	0.91866
M00079	60.6782	245.233	2.0149	7.25697	Keratan sulfate degradation	Glycosaminoglycan metabolism	0.08219
M00080	0.30048	6.12494	4.34937	1.68379	Lipopolysaccharide biosynthesis, inner core => outer core => O-antigen	Lipopolysaccharide metabolism	0.96939
M00081	0.09137	9.18182	6.65097	2.21306	Pectin degradation	Other carbohydrate metabolism	0.40528
M00082	2277.2	1804.08	-0.336	10.9948	Fatty acid biosynthesis, initiation	Fatty acid metabolism	0.00043
M00083	3659.89	1629.34	-1.1675	11.3688	Fatty acid biosynthesis, elongation	Fatty acid metabolism	0.00508
M00085	1486.3	833.557	-0.8344	10.1798	Fatty acid biosynthesis, elongation, mitochondria	Fatty acid metabolism	0.07184
M00086	830.924	1508.93	0.86074	10.1922	beta-Oxidation, acyl-CoA synthesis	Fatty acid metabolism	0.14672
M00087	2401.47	2197.97	-0.1277	11.1672	beta-Oxidation	Fatty acid metabolism	0.49673
M00088	938.679	459.859	-1.0294	9.4497	Ketone body biosynthesis, acetyl-CoA => acetoacetate/3-hydroxybutyrate/acetone	Fatty acid metabolism	0.05375
M00089	384.327	445.512	0.21313	8.69669	Triacylglycerol biosynthesis	Lipid metabolism	0.93683
M00090	28.9987	78.0768	1.4289	5.74248	Phosphatidylcholine (PC) biosynthesis, choline => PC	Lipid metabolism	0.08878
M00091	49.7442	2.50942	-4.3091	4.70746	Phosphatidylcholine (PC) biosynthesis, PE => PC	Lipid metabolism	0.14372
M00092	72.1649	136.065	0.91492	6.70203	Phosphatidylethanolamine (PE) biosynthesis, ethanolamine => PE	Lipid metabolism	0.02927
M00093	79.8099	143.917	0.85059	6.80559	Phosphatidylethanolamine (PE) biosynthesis, PA => PS => PE	Lipid metabolism	0.34468
M00094	433.613	740.431	0.77196	9.19727	Ceramide biosynthesis	Lipid metabolism	0.15977
M00095	1251.72	806.002	-0.6351	10.0068	C5 isoprenoid biosynthesis, mevalonate pathway	Terpenoid backbone biosynthesis	0.50284
M00096	1252.04	1066.23	-0.2318	10.1788	C5 isoprenoid biosynthesis, non-mevalonate pathway	Terpenoid backbone biosynthesis	0.68373
M00097	266.322	244.519	-0.1232	7.99673	beta-Carotene biosynthesis, GGAP => beta-carotene	Other terpenoid biosynthesis	0.85481
M00098	148.891	172.495	0.2123	7.32816	Acylglycerol degradation	Lipid metabolism	0.99761
M00099	489.995	785.685	0.68118	9.31705	Sphingosine biosynthesis	Lipid metabolism	0.33092
M00100	23.2893	122.056	2.3898	6.18334	Sphingosine degradation	Lipid metabolism	0.03583
M00101	838.337	594.871	-0.495	9.48503	Cholesterol biosynthesis, squalene 2,3-epoxide => cholesterol	Sterol biosynthesis	0.06065
M00102	714.443	596.695	-0.2598	9.3566	Ergocalciferol biosynthesis	Sterol biosynthesis	0.25526

M00104	52.0391	231.738	2.15483	7.14862	Bile acid biosynthesis, cholesterol => cholate/chenodeoxycholate	Sterol biosynthesis	0.11966
M00106	1.35243	0.05004	-4.7562	-0.512	Conjugated bile acid biosynthesis, cholate => taurocholate/glycocholate	Sterol biosynthesis	0.04822
M00107	0.13723	2.32207	4.08071	0.29825	Steroid hormone biosynthesis, cholesterol => prognenolone => progesterone	Sterol biosynthesis	0.86223
M00109	1.61582	4.42713	1.45411	1.59525	C21-Steroid hormone biosynthesis, progesterone => cortisol/cortisone	Sterol biosynthesis	0.77727
M00110	0.1496	4.82393	5.01104	1.31427	C19/C18-Steroid hormone biosynthesis, pregnenolone => androstenedione => estrone	Sterol biosynthesis	0.58149
M00112	97.8085	87.0043	-0.1689	6.52992	Tocopherol/tocotorienol biosynthesis	Cofactor and vitamin biosynthesis	0.73767
M00113	631.54	996.866	0.65853	9.66924	Jasmonic acid biosynthesis	Fatty acid metabolism	0.54126
M00114	2481.93	1698.6	-0.5471	11.0295	Ascorbate biosynthesis, plants, glucose-6P => ascorbate	Cofactor and vitamin biosynthesis	0.42405
M00115	166.645	183.249	0.13702	7.45077	NAD biosynthesis, aspartate => NAD	Cofactor and vitamin biosynthesis	0.63294
M00116	14.6583	32.5423	1.1506	4.56073	Menaquinone biosynthesis, chorismate => menaquinone	Cofactor and vitamin biosynthesis	0.67325
M00117	17.8928	34.908	0.96418	4.72249	Ubiquinone biosynthesis, prokaryotes, chorismate => ubiquinone	Cofactor and vitamin biosynthesis	0.61788
M00118	167.63	192.946	0.20292	7.49416	Glutathione biosynthesis, glutamate => glutathione	Cofactor and vitamin biosynthesis	0.96043
M00119	814.375	780.601	-0.0611	9.63932	Pantothenate biosynthesis, valine/L-aspartate => pantothenate	Cofactor and vitamin biosynthesis	0.60533
M00120	22.4923	45.9944	1.03203	5.09775	Coenzyme A biosynthesis, pantothenate => CoA	Cofactor and vitamin biosynthesis	0.66998
M00121	5362.33	2154.68	-1.3154	11.8759	Heme biosynthesis, glutamate => protoheme/siroheme	Cofactor and vitamin biosynthesis	0.0121
M00122	6.3867	14.7948	1.21194	3.40473	Cobalamin biosynthesis, cobinamide => cobalamin	Cofactor and vitamin biosynthesis	0.76842
M00123	212.804	98.8885	-1.1056	7.28398	Biotin biosynthesis, pimeloyl-ACP/CoA => biotin	Cofactor and vitamin biosynthesis	0.55826
M00124	791.537	1079.69	0.44789	9.86977	Pyridoxal biosynthesis, erythrose-4P => pyridoxal-5P	Cofactor and vitamin biosynthesis	0.60495
M00125	192.158	304.359	0.66348	7.9557	Riboflavin biosynthesis, GTP => riboflavin/FMN/FAD	Cofactor and vitamin biosynthesis	0.34771
M00126	145.134	592.826	2.03022	8.5274	Tetrahydrofolate biosynthesis, GTP => THF	Cofactor and vitamin biosynthesis	0.16327
M00127	54.8462	19.8691	-1.4649	5.22333	Thiamine biosynthesis, AIR => thiamine-P/thiamine-2P	Cofactor and vitamin biosynthesis	0.1073
M00128	15.1862	50.3699	1.7298	5.03466	Ubiquinone biosynthesis, eukaryotes, 4-hydroxybenzoate => ubiquinone	Cofactor and vitamin biosynthesis	0.45719
M00129	6112.69	2321.11	-1.397	12.042	Ascorbate biosynthesis, animals, glucose-1P => ascorbate	Cofactor and vitamin biosynthesis	0.00712
M00130	55.6352	278.551	2.32387	7.38451	Inositol phosphate metabolism, PI=> PIP2 => Ins(1,4,5)P3 => Ins(1,3,4,5)P4	Lipid metabolism	0.00723
M00131	36.1291	89.9995	1.31676	5.97875	Inositol phosphate metabolism, Ins(1,3,4,5)P4 => Ins(1,3,4)P3 => myo-inositol	Lipid metabolism	0.66641
M00132	6.561	52.3227	2.99545	4.8798	Inositol phosphate metabolism, Ins(1,3,4)P3 => phytate	Lipid metabolism	0.32329
M00133	1319.12	942.11	-0.4856	10.1429	Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine	Polyamine biosynthesis	0.14415
M00134	1179.53	635.459	-0.8923	9.82575	Polyamine biosynthesis, arginine => ornithine => putrescine	Polyamine biosynthesis	0.15525
M00135	687.075	520.608	-0.4003	9.23803	GABA biosynthesis, eukaryotes, putrescine => GABA	Polyamine biosynthesis	0.59535
M00136	0.9583	0.26049	-1.8792	-0.7146	GABA biosynthesis, prokaryotes, putrescine => GABA	Polyamine biosynthesis	0.59716
M00137	4.70302	2.86997	-0.7126	1.92086	Flavanone biosynthesis, phenylalanine => naringenin	Biosynthesis of secondary metabolites	0.19914
M00138	1.01682	1.32217	0.37883	0.22589	Flavonoid biosynthesis, naringenin => pelargonidin	Biosynthesis of secondary metabolites	0.85974
M00140	7415.64	7088.5	-0.0651	12.8242	C1-unit interconversion, prokaryotes	Cofactor and vitamin biosynthesis	0.97004
M00141	6930	6861.64	-0.0143	12.7515	C1-unit interconversion, eukaryotes	Cofactor and vitamin	0.93004

						biosynthesis	
M00142	3307.5	1128.35	-1.5515	11.115	NADH:ubiquinone oxidoreductase, mitochondria	ATP synthesis	0.00622
M00143	2276.55	4151.38	0.86675	11.6501	NADH dehydrogenase (ubiquinone) Fe-S protein/flavoprotein complex, mitochondria	ATP synthesis	0.71354
M00144	296.543	257.933	-0.2012	8.11498	NADH:quinone oxidoreductase, prokaryotes	ATP synthesis	0.09525
M00145	4.12588	140.359	5.08827	6.17477	NAD(P)H:quinone oxidoreductase, chloroplasts and cyanobacteria	ATP synthesis	0.34274
M00146	1089.48	678.693	-0.6828	9.78804	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex	ATP synthesis	0.03557
M00147	167.966	132.001	-0.3476	7.22866	NADH dehydrogenase (ubiquinone) 1 beta subcomplex	ATP synthesis	0.18916
M00148	3122.96	4408.1	0.49724	11.8786	Succinate dehydrogenase (ubiquinone)	ATP synthesis	0.28426
M00149	374.21	786.415	1.07144	9.18069	Succinate dehydrogenase, prokaryotes	ATP synthesis	0.81485
M00150	3.67537	18.2578	2.31255	3.45504	Fumarate reductase, prokaryotes	ATP synthesis	0.27897
M00151	10940	20823.1	0.92857	13.9551	Cytochrome bc1 complex respiratory unit	ATP synthesis	0.41901
M00152	11641.2	21287.4	0.87076	14.0071	Cytochrome bc1 complex	ATP synthesis	0.25289
M00153	0.37566	0.61282	0.70604	-1.0167	Cytochrome d ubiquinol oxidase	ATP synthesis	0.98543
M00154	11644.5	13952.7	0.2609	13.6437	Cytochrome c oxidase	ATP synthesis	0.45881
M00155	374.55	470.285	0.32838	8.72253	Cytochrome c oxidase, prokaryotes	ATP synthesis	0.82323
M00156	0.0346	1.10859	5.00169	-0.8069	Cytochrome c oxidase, cbb3-type	ATP synthesis	0.73021
M00157	14309.3	8316.94	-0.7828	13.4657	F-type ATPase, prokaryotes and chloroplasts	ATP synthesis	0.13765
M00158	61795.3	41786.8	-0.5644	15.6604	F-type ATPase, eukaryotes	ATP synthesis	0.08471
M00159	1.51378	6.74103	2.15481	2.04524	V/A-type ATPase, prokaryotes	ATP synthesis	0.39198
M00160	7678.96	4816.95	-0.6728	12.6092	V-type ATPase, eukaryotes	ATP synthesis	0.29722
M00161	51818.7	27798	-0.8985	15.2808	Photosystem II	Photosynthesis	0.02385
M00162	6995.23	3608.06	-0.9551	12.3722	Cytochrome b6f complex	ATP synthesis	0.00089
M00163	13081.4	4314.49	-1.6003	13.0865	Photosystem I	Photosynthesis	0.00039
M00165	448564	133435	-1.7492	18.1507	Reductive pentose phosphate cycle (Calvin cycle)	Carbon fixation	8.07E-24
M00166	324198	96223.2	-1.7524	17.6815	Reductive pentose phosphate cycle, ribulose-5P => glyceraldehyde-3P	Carbon fixation	1.81E-16
M00167	124366	37211.8	-1.7408	16.3019	Reductive pentose phosphate cycle, glyceraldehyde-3P => ribulose-5P	Carbon fixation	1.29E-09
M00168	12190.5	4655.01	-1.3889	13.0401	CAM (Crassulacean acid metabolism), dark	Carbon fixation	0.00344
M00169	1471.05	9767.6	2.73116	12.4562	CAM (Crassulacean acid metabolism), light	Carbon fixation	2.43E-05
M00170	2571.74	2503.26	-0.0389	11.3092	C4-dicarboxylic acid cycle, phosphoenolpyruvate carboxykinase type	Carbon fixation	0.01603
M00171	14792.7	15060.5	0.02588	13.8656	C4-dicarboxylic acid cycle, NAD - malic enzyme type	Carbon fixation	9.42E-11
M00172	2043.21	10679.4	2.38592	12.6351	C4-dicarboxylic acid cycle, NADP - malic enzyme type	Carbon fixation	0.00011
M00173	10242.1	19086.1	0.89801	13.84	Reductive citrate cycle (Arnon-Buchanan cycle)	Carbon fixation	0.33036
M00174	0.55413	0.03606	-3.9416	-1.7607	Methane oxidation, methanotroph, methane => formaldehyde	Methane metabolism	0.17208
M00176	2616.53	2521.55	-0.0533	11.327	Assimilatory sulfate reduction, sulfate => H2S	Sulfur metabolism	0.24312
M00177	527994	214885	-1.297	18.5028	Ribosome, eukaryotes	Ribosome	4.52E-54
M00178	28629.4	23080.2	-0.3108	14.6581	Ribosome, bacteria	Ribosome	3.68E-05
M00179	334187	148979	-1.1655	17.8822	Ribosome, archaea	Ribosome	2.67E-24

M00180	1316.72	1866.98	0.50376	10.6365	RNA polymerase II, eukaryotes	RNA polymerase	0.02906
M00181	1063.24	1213.7	0.19094	10.1529	RNA polymerase III, eukaryotes	RNA polymerase	0.32021
M00182	1076.55	1240.39	0.20437	10.178	RNA polymerase I, eukaryotes	RNA polymerase	0.47041
M00183	566.614	1140.26	1.00893	9.73714	RNA polymerase, bacteria	RNA polymerase	0.19638
M00184	3.52214	20.1099	2.51338	3.56267	RNA polymerase, archaea	RNA polymerase	0.91256
M00185	3.78543	2.16399	-0.8068	1.57275	Sulfate transport system	Mineral and organic ion transport system	0.95009
M00186	0.06806	0.5622	3.04617	-1.666	Tungstate transport system	Mineral and organic ion transport system	0.89145
M00188	0.68934	8.32345	3.59389	2.17197	NitT/TauT family transport system	Mineral and organic ion transport system	0.51127
M00189	0.03149	1.32346	5.39312	-0.5618	Molybdate transport system	Mineral and organic ion transport system	0.34683
M00190	0.75856	2.74121	1.85348	0.80726	Iron(III) transport system	Mineral and organic ion transport system	0.87174
M00191	0.53099	0.04197	-3.6614	-1.8035	Thiamine transport system	Mineral and organic ion transport system	0.16526
M00192	0.98156	0.01484	-6.0477	-1.0052	Putative thiamine transport system	Mineral and organic ion transport system	0.08015
M00193	23.0714	2.18179	-3.4025	3.6584	Putative spermidine/putrescine transport system	Mineral and organic ion transport system	0.33014
M00194	2.61556	2.61556	0	1.38712	Maltose/maltodextrin transport system	Saccharide, polyol, and lipid transport system	1
M00196	2.65734	3.46833	0.38425	1.61487	Raffinose/stachyose/melibiose transport system	Saccharide, polyol, and lipid transport system	1
M00197	2.61556	2.61556	0	1.38712	Putative fructooligosaccharide transport system	Saccharide, polyol, and lipid transport system	1
M00198	0.11532	9.70325	6.39475	2.29551	Putative sn-glycerol-phosphate transport system	Saccharide, polyol, and lipid transport system	0.46722
M00200	2.61556	2.61556	0	1.38712	Putative sorbitol/mannitol transport system	Saccharide, polyol, and lipid transport system	1
M00201	4.43052	3.71884	-0.2526	2.02669	alpha-Glucoside transport system	Saccharide, polyol, and lipid transport system	1
M00204	0.43472	1.12098	1.36661	-0.3624	Trehalose/maltose transport system	Saccharide, polyol, and lipid transport system	0.86035
M00206	2.61556	2.61556	0	1.38712	Cellobiose transport system	Saccharide, polyol, and lipid transport system	1
M00207	4.90482	16.4029	1.74168	3.41331	Putative multiple sugar transport system	Saccharide, polyol, and lipid transport system	1
M00208	0.02157	16.9118	9.61471	3.0818	Glycine betaine/proline transport system	Mineral and organic ion transport system	0.05347
M00210	5.38999	6.73085	0.32051	2.59942	Phospholipid transport system	Saccharide, polyol, and lipid transport system	0.98182
M00211	0.38319	3.23264	3.0766	0.85433	Putative ABC transport system	Saccharide, polyol, and lipid transport system	0.57308
M00212	4.19272	0.84762	-2.3064	1.33352	Ribose transport system	Saccharide, polyol, and lipid transport system	0.69885
M00215	0.56667	0.11237	-2.3343	-1.5584	D-Xylose transport system	Saccharide, polyol, and lipid transport system	0.36212
M00218	0.33514	0.87585	1.38593	-0.7238	Fructose transport system	Saccharide, polyol, and lipid transport system	0.91038
M00220	0.30744	2.43319	2.98449	0.45451	Rhamnose transport system	Saccharide, polyol, and lipid transport	0.39616

						system	
M00221	0.33556	6.92499	4.36719	1.86008	Putative simple sugar transport system	Saccharide, polyol, and lipid transport system	0.76697
M00222	255.917	40.3041	-2.6667	7.21053	Phosphate transport system	Phosphate and amino acid transport system	0.26562
M00223	0.02538	2.66753	6.71589	0.42916	Phosphonate transport system	Phosphate and amino acid transport system	0.3243
M00225	0.05858	0.63235	3.43215	-1.5334	Lysine/arginine/ornithine transport system	Phosphate and amino acid transport system	0.86737
M00227	0.03743	0.99341	4.73003	-0.9562	Glutamine transport system	Phosphate and amino acid transport system	0.76681
M00229	0.04826	0.74259	3.94369	-1.3385	Arginine transport system	Phosphate and amino acid transport system	0.83111
M00230	1.88562	2.88285	0.61246	1.25353	Glutamate/aspartate transport system	Phosphate and amino acid transport system	0.67083
M00231	0.02672	2.13819	6.32242	0.11431	Octopine/nopaline transport system	Phosphate and amino acid transport system	0.56481
M00232	9.67435	19.4777	1.00959	3.86553	General L-amino acid transport system	Phosphate and amino acid transport system	0.69623
M00234	0.27552	1.20526	2.12912	-0.4336	Cystine transport system	Phosphate and amino acid transport system	0.94707
M00235	0.04117	0.86803	4.39818	-1.1373	Arginine/ornithine transport system	Phosphate and amino acid transport system	0.80172
M00236	11.1949	19.8077	0.82322	3.95432	Putative polar amino acid transport system	Phosphate and amino acid transport system	0.53593
M00237	21.3763	18.4827	-0.2098	4.31684	Branched-chain amino acid transport system	Phosphate and amino acid transport system	0.99198
M00238	0.72195	0.52537	-0.4586	-0.6812	D-Methionine transport system	Phosphate and amino acid transport system	0.72107
M00239	6.00809	7.21505	0.2641	2.72499	Peptides/nickel transport system	Peptide and nickel transport system	0.99814
M00240	2.03602	1.20293	-0.7592	0.69552	Iron complex transport system	Metallic cation, iron-siderophore and vitamin B12 transport system	0.62066
M00242	0.06388	2.76959	5.4381	0.50257	Zinc transport system	Metallic cation, iron-siderophore and vitamin B12 transport system	0.33589
M00243	0.03908	0.93433	4.57944	-1.0389	Manganese/iron transport system	Metallic cation, iron-siderophore and vitamin B12 transport system	0.49548
M00244	5.13034	3.42852	-0.5815	2.09742	Putative zinc/manganese transport system	Metallic cation, iron-siderophore and vitamin B12 transport system	0.77764
M00250	0.10076	0.45053	2.16074	-1.8591	Lipopolysaccharide transport system	ABC-2 type and other transport systems	0.93854
M00252	0.23778	0.23778	0	-2.0723	Lipooligosaccharide transport system	ABC-2 type and other transport systems	0.93854
M00253	0.66191	0.751	0.18219	-0.5013	Sodium transport system	ABC-2 type and other transport systems	0.46954
M00254	2.53712	8.8394	1.80076	2.50799	ABC-2 type transport system	ABC-2 type and other transport systems	0.40249
M00255	0.17666	1.01689	2.52515	-0.7447	Lipoprotein-releasing system	ABC-2 type and other transport systems	0.94651
M00256	0.23778	0.23778	0	-2.0723	Cell division transport system	ABC-2 type and other transport systems	0.93854
M00258	11.412	12.5044	0.13188	3.57992	Putative ABC transport system	ABC-2 type and other transport systems	0.81277
M00259	0.09692	0.46133	2.25099	-1.841	Heme transport system	ABC-2 type and other transport systems	0.93854
M00260	7.22558	12.6693	0.81015	3.31432	DNA polymerase III complex, bacteria	DNA polymerase	0.55704
M00261	104.034	316.354	1.60448	7.71558	DNA polymerase alpha / primase complex	DNA polymerase	0.04107
M00262	66.4178	162.147	1.28766	6.83646	DNA polymerase delta complex	DNA polymerase	0.70195

M00263	30.5909	133.02	2.12047	6.35413	DNA polymerase epsilon complex	DNA polymerase	0.06709
M00273	0.09147	0.47554	2.37821	-1.8186	PTS system, fructose-specific II component	Phosphotransferase system (PTS)	0.93854
M00284	16.073	65.2389	2.0211	5.34539	Origin recognition complex	Replication system	0.09452
M00285	487.294	994.035	1.0285	9.53268	MCM complex	Replication system	0.0408
M00286	55.4237	86.409	0.64068	6.14805	GINS complex	Replication system	0.50561
M00288	385.511	965.562	1.3246	9.39989	RPA complex	Replication system	0.2054
M00289	230.158	841.295	1.86999	9.06535	RF-C complex	Replication system	0.00041
M00290	32.63	184.846	2.50206	6.76471	Holo-TFIIH complex	Repair system	0.00139
M00291	15.4505	209.399	3.76052	6.81282	MRN complex	Repair system	0.00544
M00292	13.1045	208.127	3.98933	6.78941	MRX complex	Repair system	0.00063
M00293	0.0512	5.92652	6.85494	1.57959	DNA polymerase zeta complex	DNA polymerase	0.11555
M00294	0.02152	18.4704	9.74521	3.20883	DNA polymerase gamma complex	DNA polymerase	0.13257
M00295	1044.46	2455.38	1.23319	10.7731	BRCA1-associated genome surveillance complex (BASC)	Repair system	1.00E-06
M00296	86.3536	147.008	0.76757	6.86643	BER complex	Repair system	0.83877
M00297	13.8843	32.7346	1.23736	4.54284	DNA-PK complex	Repair system	0.44947
M00299	0.07783	1.91606	4.62165	-0.0044	Spermidine/putrescine transport system	Mineral and organic ion transport system	0.6481
M00300	0.23778	0.23778	0	-2.0723	Putrescine transport system	Mineral and organic ion transport system	0.93854
M00307	5720.98	4015.05	-0.5108	12.2491	Pyruvate oxidation, pyruvate => acetyl-CoA	Central carbohydrate metabolism	0.09233
M00308	180973	61353.7	-1.5606	16.8866	Semi-phosphorylative Entner-Doudoroff pathway, gluconate => glycerate-3P	Central carbohydrate metabolism	1.84E-08
M00309	0.02632	1.92007	6.18866	-0.0392	Non-phosphorylative Entner-Doudoroff pathway, gluconate/galactonate => glycerate	Central carbohydrate metabolism	0.52715
M00319	0.0466	0.70737	3.92415	-1.4074	Manganese/zinc/iron transport system	Metallic cation, iron-siderophore and vitamin B12 transport system	0.89767
M00320	1.23097	0.09995	-3.6225	-0.5876	Lipopolysaccharide export system	ABC-2 type and other transport systems	0.28564
M00323	1.90914	24.8001	3.69935	3.73926	Urea transport system	Phosphate and amino acid transport system	0.08751
M00324	0.03311	1.18819	5.16541	-0.7116	Dipeptide transport system	Peptide and nickel transport system	0.26786
M00325	1.11134	52.1109	5.55121	4.73396	alpha-Hemolysin/cyclolysin transport system	Bacterial secretion system	0.00853
M00326	0.594	52.1286	6.45546	4.72035	RTX toxin transport system	Bacterial secretion system	0.02399
M00330	0.03544	1.07051	4.91674	-0.8547	Adhesin protein transport system	Bacterial secretion system	0.45985
M00331	1.24219	4.29861	1.79098	1.4701	Type II general secretion pathway	Bacterial secretion system	0.84158
M00332	0.10312	3.36307	5.02732	0.79335	Type III secretion system	Bacterial secretion system	0.85511
M00333	0.23778	0.23778	0	-2.0723	Type IV secretion system	Bacterial secretion system	0.93854
M00334	4.76745	41.1247	3.10872	4.52017	Type VI secretion system	Bacterial secretion system	0.15797
M00335	1126.45	1050.78	-0.1003	10.0883	Sec (secretion) system	Bacterial secretion system	0.04717
M00336	129.26	77.9283	-0.7301	6.6948	Twin-arginine translocation (Tat) system	Bacterial secretion system	0.33916
M00337	5492.91	5480.8	-0.0032	12.4218	Immunoproteasome	Proteasome	0.99977
M00338	357.473	547.426	0.61483	8.82161	Cysteine biosynthesis, homocysteine + serine => cysteine	Cysteine and methionine metabolism	0.42966

M00339	0.29039	52.4464	7.49669	4.72074	RaxAB-RaxC type I secretion system	Bacterial secretion system	0.00245
M00340	6211.41	6207.16	-0.001	12.6002	Proteasome, 20S core particle	Proteasome	1
M00341	3016.86	3729.8	0.30605	11.72	Proteasome, 19S regulatory particle (PA700)	Proteasome	0.96184
M00342	3.16467	1.5545	-1.0256	1.23853	Bacterial proteasome	Proteasome	0.7463
M00343	3.2195	2.2132	-0.5407	1.44167	Archaeal proteasome	Proteasome	0.75832
M00344	33175.3	21650.2	-0.6157	14.7426	Formaldehyde assimilation, xylulose monophosphate pathway	Methane metabolism	0.06145
M00345	23326.2	19337.4	-0.2706	14.3807	Formaldehyde assimilation, ribulose monophosphate pathway	Methane metabolism	0.0506
M00346	21170.8	14090.8	-0.5873	14.1058	Formaldehyde assimilation, serine pathway	Methane metabolism	0.30827
M00350	24.5562	257.673	3.39138	7.14072	Capsaicin biosynthesis, L-Phenylalanine => Capsaicin	Aromatic amino acid metabolism	0.08537
M00351	4191.28	4717.65	0.17068	12.121	Spliceosome, U1-snRNP	Spliceosome	0.99999
M00352	4595.49	5329.77	0.21385	12.2769	Spliceosome, U2-snRNP	Spliceosome	0.99989
M00353	24636.8	23729.3	-0.0541	14.5617	Spliceosome, Prp19/CDC5L complex	Spliceosome	0.70257
M00354	5926.44	6827.11	0.20411	12.6386	Spliceosome, U4/U6.U5 tri-snRNP	Spliceosome	0.99999
M00355	29693.4	29671.5	-0.0011	14.8573	Spliceosome, 35S U5-snRNP	Spliceosome	0.99983
M00356	1.2607	3.04715	1.27324	1.10697	Methanogenesis, methanol => methane	Methane metabolism	0.99997
M00357	189.474	392.89	1.05213	8.18578	Methanogenesis, acetate => methane	Methane metabolism	0.99754
M00358	0.7719	3.60251	2.22251	1.12909	Coenzyme M biosynthesis	Methane metabolism	0.07817
M00359	9958.27	7095.79	-0.4889	13.0578	Aminoacyl-tRNA biosynthesis, eukaryotes	Aminoacyl tRNA	0.07305
M00360	9891.79	6957.26	-0.5077	13.0404	Aminoacyl-tRNA biosynthesis, prokaryotes	Aminoacyl tRNA	0.34653
M00361	7073.03	3092.87	-1.1934	12.3115	Nucleotide sugar biosynthesis, eukaryotes	Nucleotide sugar	0.00207
M00362	7049.39	3066.85	-1.2007	12.3044	Nucleotide sugar biosynthesis, prokaryotes	Nucleotide sugar	0.00105
M00364	671.246	642.583	-0.063	9.35956	C10-C20 isoprenoid biosynthesis, bacteria	Terpenoid backbone biosynthesis	0.97725
M00365	176.508	230.824	0.38706	7.67006	C10-C20 isoprenoid biosynthesis, archaea	Terpenoid backbone biosynthesis	0.96202
M00366	1026.2	793.239	-0.3715	9.82928	C10-C20 isoprenoid biosynthesis, plants	Terpenoid backbone biosynthesis	0.7135
M00367	537.524	390.934	-0.4594	8.85869	C10-C20 isoprenoid biosynthesis, non-plant eukaryotes	Terpenoid backbone biosynthesis	0.62167
M00368	21307.8	13026.1	-0.71	14.0673	Ethylene biosynthesis, methionine => ethylene	Cysteine and methionine metabolism	0.09953
M00371	0.13403	2.59585	4.27559	0.44884	Castasterone biosynthesis, campesterol => castasterone	Other terpenoid biosynthesis	0.74319
M00372	192.787	496.909	1.36597	8.42982	Abscisic acid biosynthesis, beta-carotene => abscisic acid	Other terpenoid biosynthesis	0.72173
M00373	1211.04	881.341	-0.4585	10.0309	Ethylmalonyl pathway	Other carbohydrate metabolism	0.25717
M00374	2910.4	3973.15	0.44907	11.7489	Dicarboxylate-hydroxybutyrate cycle	Carbon fixation	0.58676
M00375	882.592	410.788	-1.1034	9.33693	Hydroxypropionate-hydroxybutylate cycle	Carbon fixation	0.98159
M00376	1362.14	1545.91	0.18258	10.5058	3-Hydroxypropionate bi-cycle	Carbon fixation	0.127
M00377	642.322	393.506	-0.7069	9.01657	Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)	Carbon fixation	0.69872
M00378	0.03099	1.33142	5.42511	-0.5538	F420 biosynthesis	Methane metabolism	0.36128
M00379	7410.73	8397.72	0.18038	12.9484	SCF-MET30 complex	Ubiquitin system	0.99956
M00380	7412.21	8401.93	0.18082	12.9489	SCF-BTRC complex	Ubiquitin system	0.99969
M00381	7410.49	8398.04	0.18048	12.9484	SCF-SKP2 complex	Ubiquitin system	0.99955
M00382	7410.7	8394.96	0.17991	12.9482	SCF-FBS complex	Ubiquitin system	0.9999

M00383	1205.35	1124.13	-0.1006	10.1858	ECV complex	Ubiquitin system	0.99938
M00384	1118.87	1159.1	0.05096	10.1535	Cul3-SPOP complex	Ubiquitin system	0.99521
M00385	1153.84	1160.35	0.00812	10.1763	Cul4-DDB1-DDB2 complex	Ubiquitin system	0.99919
M00386	1157.06	1163.89	0.0085	10.1805	Cul4-DDB1-CSA complex	Ubiquitin system	0.99919
M00387	7412.5	8400.84	0.18057	12.9489	SCF-FBW7 complex	Ubiquitin system	0.99974
M00388	143.172	99.4316	-0.526	6.92246	ECS complex	Ubiquitin system	0.93921
M00389	118.031	247.278	1.06697	7.51298	APC/C complex	Ubiquitin system	0.14051
M00390	68.8831	172.315	1.32283	6.91408	Exosome, archaea	RNA processing	0.30849
M00391	104.518	281.087	1.42727	7.59098	Exosome, eukaryotes	RNA processing	0.36862
M00392	45.7594	56.5753	0.3061	5.67715	Ski complex	RNA processing	0.58026
M00393	40.3816	99.6429	1.30307	6.12954	TRAMP complex	RNA processing	0.72638
M00394	12707.6	5676.63	-1.1626	13.1662	RNA degradosome	RNA processing	0.00212
M00395	511.96	505.023	-0.0197	8.99008	Decapping complex	RNA processing	0.39599
M00396	576.444	802.934	0.4781	9.4298	Lsm 2-8 complex	Spliceosome	0.95499
M00397	562.267	789.287	0.4893	9.4004	Lsm 1-7 complex	Spliceosome	0.94881
M00398	4134.61	4576.34	0.14645	12.0886	Sm core complex	Spliceosome	1
M00399	59.7198	96.5994	0.6938	6.28835	Cap binding complex	Spliceosome	0.60179
M00400	3760.75	2924.42	-0.3629	11.7067	p97-Ufd1-Npl4 complex	Protein processing	0.94783
M00401	1567.83	1056.9	-0.5689	10.358	Sec61 complex	Protein processing	0.31619
M00402	547.912	325.085	-0.7531	8.76983	Translocon-associated protein (TRAP) complex	Protein processing	0.36119
M00403	3862.1	3124.68	-0.3057	11.7704	HRD1/SEL1 ERAD complex	Protein processing	0.68512
M00404	913.947	1206.49	0.40063	10.0501	COPII complex	Protein processing	0.70012
M00405	100.788	144.254	0.51728	6.93689	THC complex	RNA processing	0.84746
M00406	2362.02	2262.99	-0.0618	11.1752	TREX complex	RNA processing	0.81888
M00407	7414.01	8397.27	0.17967	12.9487	SCF-CDC4 complex	Ubiquitin system	0.9998
M00408	7.51366	3.69692	-1.0232	2.48679	ESCRT-0 complex	Protein processing	0.51392
M00409	144.521	88.3294	-0.7103	6.86326	ESCRT-I complex	Protein processing	0.29547
M00410	15.1563	35.863	1.24258	4.67297	ESCRT-II complex	Protein processing	0.39312
M00411	7412.8	8395.92	0.17967	12.9484	SCF-GRR1 complex	Ubiquitin system	0.99975
M00412	905.603	853.397	-0.0857	9.78054	ESCRT-III complex	Protein processing	0.22675
M00413	2.08195	7.45687	1.84064	2.25381	FA core complex	Repair system	0.20509
M00414	37.5841	99.3321	1.40214	6.09715	Bloom's syndrome complex	Repair system	0.50694
M00415	473.953	320.978	-0.5623	8.63469	Fatty acid biosynthesis, elongation, endoplasmic reticulum	Fatty acid metabolism	0.679
M00416	1.16695	6.23883	2.41853	1.88865	Cytochrome aa3-600 menaquinol oxidase	ATP synthesis	0.33007
M00417	0.87568	0.05042	-4.1184	-1.1108	Cytochrome o ubiquinol oxidase	ATP synthesis	0.10571
M00419	0.41509	0.38243	-0.1182	-1.3264	Cymene degradation, p-cymene => p-cumate	Aromatics degradation	0.83744
M00424	44.7963	11.1405	-2.0076	4.80572	Shelterin complex	Replication system	0.41564
M00425	505.474	581.464	0.20205	9.08605	H/ACA ribonucleoprotein complex	RNA processing	0.94459
M00426	40.6402	46.204	0.18511	5.44036	Survival motor neuron (SMN) complex	RNA processing	0.92637
M00427	791.039	1283.56	0.69833	10.0186	Nuclear pore complex	RNA processing	0.45929
M00428	3890.62	2554.81	-0.6068	11.6541	eIF4F complex	RNA processing	0.55334
M00429	7.68053	26.4522	1.78411	4.09308	Competence-related DNA transformation transporter	Bacterial secretion system	0.38686

M00430	2840.47	2946.32	0.05278	11.4985	Exon junction complex (EJC)	RNA processing	0.14588
M00432	1266.73	293.568	-2.1093	9.60761	Leucine biosynthesis, 2-oxoisovalerate => 2-oxoisocaproate	Branched-chain amino acid metabolism	8.24E-05
M00433	10.5266	8.50763	-0.3072	3.25053	Lysine biosynthesis, 2-oxoglutarate => 2-oxoadipate	Lysine metabolism	0.99808
M00434	0.36562	2.85224	2.96367	0.6861	PhoR-PhoB (phosphate starvation response) two-component regulatory system	Two-component regulatory system	0.73274
M00435	0.12056	0.40695	1.75505	-1.9227	Taurine transport system	Mineral and organic ion transport system	0.93854
M00436	0.28046	1.06684	1.92751	-0.5699	Sulfonate transport system	Mineral and organic ion transport system	0.91849
M00438	0.22378	3.52495	3.97742	0.9064	Nitrate/nitrite transport system	Mineral and organic ion transport system	0.79622
M00439	0.31657	0.76011	1.26369	-0.8934	Oligopeptide transport system	Peptide and nickel transport system	0.97886
M00440	0.10749	0.43474	2.01597	-1.883	Nickel transport system	Peptide and nickel transport system	0.93854
M00444	0.19511	7.52526	5.26935	1.94867	PhoQ-PhoP (magnesium transport) two-component regulatory system	Two-component regulatory system	0.63992
M00445	0.27976	1.41177	2.33525	-0.2417	EnvZ-OmpR (osmotic stress response) two-component regulatory system	Two-component regulatory system	0.99624
M00452	0.34311	2.09679	2.61146	0.28682	CusS-CusR (copper tolerance) two-component regulatory system	Two-component regulatory system	0.99806
M00454	0.23778	0.23778	0	-2.0723	KdpD-KdpE (potassium transport) two-component regulatory system	Two-component regulatory system	0.93854
M00455	0.74039	0.48651	-0.6058	-0.705	TorS-TorR (TMAO respiration) two-component regulatory system	Two-component regulatory system	0.61065
M00456	0.23778	0.23778	0	-2.0723	ArcB-ArcA (anoxic redox control) two-component regulatory system	Two-component regulatory system	0.93854
M00458	0.7441	0.02174	-5.0968	-1.3849	ResE-ResD (aerobic and anaerobic respiration) two-component regulatory system	Two-component regulatory system	0.11197
M00459	0.16958	8.45252	5.63938	2.10804	VicK-VicR (cell wall metabolism) two-component regulatory system	Two-component regulatory system	0.62805
M00467	0.76837	0.77864	0.01915	-0.3705	SasA-RpaAB (circadian timing mediating) two-component regulatory system	Two-component regulatory system	0.89988
M00472	0.07548	0.53385	2.82219	-1.7147	NarQ-NarP (nitrate respiration) two-component regulatory system	Two-component regulatory system	0.90142
M00473	0.09748	0.45868	2.23425	-1.8464	UhpB-UhpA (hexose phosphates uptake) two-component regulatory system	Two-component regulatory system	0.93854
M00474	0.33205	2.44135	2.87822	0.47165	RcsC-RcsD-RcsB (capsule synthesis) two-component regulatory system	Two-component regulatory system	0.62268
M00475	0.23778	0.23778	0	-2.0723	BarA-UvrY (central carbon metabolism) two-component regulatory system	Two-component regulatory system	0.93854
M00478	0.05065	0.70108	3.79088	-1.4117	DegS-DegU (multicellular behavior control) two-component regulatory system	Two-component regulatory system	0.85081
M00479	0.23778	0.23778	0	-2.0723	DesK-DesR (membrane lipid fluidity regulation) two-component regulatory system	Two-component regulatory system	0.93854
M00483	0.03908	0.84323	4.43131	-1.1806	NreB-NreC (dissimilatory nitrate/nitrite reduction) two-component regulatory system	Two-component regulatory system	0.86981
M00485	0.03031	1.46382	5.59388	-0.4207	KinABCDE-Spo0FA (sporulation control) two-component regulatory system	Two-component regulatory system	0.66293
M00491	2.61556	2.61556	0	1.38712	arabinogalactan oligomer/maltooligosaccharide transport system	Saccharide, polyol, and lipid transport system	1
M00497	0.0374	0.97775	4.70828	-0.9783	GlnL-GlnG (nitrogen regulation) two-component regulatory system	Two-component regulatory system	0.47386
M00498	0.04807	0.73951	3.9435	-1.3445	NtrY-NtrX (nitrogen regulation) two-component regulatory system	Two-component regulatory system	0.83139
M00501	0.02166	14.8356	9.42013	2.8931	PilS-PilR (type 4 fimbriae synthesis) two-component regulatory system	Two-component regulatory system	0.16228

M00506	6.01222	7.08529	0.23693	2.71122	CheA-CheYBV (chemotaxis) two-component regulatory system	Two-component regulatory system	0.8626
M00507	0.68098	2.81994	2.04999	0.80773	ChpA-ChpB/PilGH (chemosensory) two-component regulatory system	Two-component regulatory system	0.21173
M00511	0.82678	1.33741	0.69387	0.11383	PleC-PleD (cell fate control) two-component regulatory system	Two-component regulatory system	0.97575
M00512	0.57627	0.92545	0.6834	-0.4134	CckA-CtrA/CpdR (cell cycle control) two-component regulatory system	Two-component regulatory system	0.99623
M00513	0.11374	0.42033	1.88578	-1.9049	LuxQN/CqsS-LuxU-LuxO (quorum sensing) two-component regulatory system	Two-component regulatory system	0.93854
M00515	0.28584	0.97729	1.77356	-0.663	FlrB-FlrC (polar flagellar synthesis) two-component regulatory system	Two-component regulatory system	0.9616
M00516	3.32545	1.0383	-1.6793	1.12557	SLN1-YPD1-SSK1/SKN7 (osmosensing) two-component regulatory system	Two-component regulatory system	0.27264
M00519	0.16215	1.0391	2.67996	-0.7355	YesM-YesN two-component regulatory system	Two-component regulatory system	0.97579
M00520	0.79814	8.37223	3.3909	2.19698	ChvG-ChvI (acidity sensing) two-component regulatory system	Two-component regulatory system	0.27752
M00523	0.75195	0.01878	-5.3233	-1.3757	RegB-RegA (redox response) two-component regulatory system	Two-component regulatory system	0.12731
M00524	2.08165	0.3307	-2.6541	0.27044	FixL-FixJ (nitrogen fixation) two-component regulatory system	Two-component regulatory system	0.26318
M00525	3726.6	1371.13	-1.4425	11.3156	Lysine biosynthesis, acetyl-DAP pathway, aspartate => lysine	Lysine metabolism	0.00018
M00526	4086.18	1858	-1.137	11.5373	Lysine biosynthesis, DAP dehydrogenase pathway, aspartate => lysine	Lysine metabolism	0.05256
M00527	4508.8	2076.03	-1.1189	11.6849	Lysine biosynthesis, DAP aminotransferase pathway, aspartate => lysine	Lysine metabolism	0.06826
M00529	2.68246	1.15309	-1.2181	0.93943	Denitrification, nitrate => nitrogen	Nitrogen metabolism	0.33417
M00530	37.5107	156.259	2.05856	6.5982	Dissimilatory nitrate reduction, nitrate => ammonia	Nitrogen metabolism	0.0452
M00531	170.816	1593.53	3.22171	9.78492	Assimilatory nitrate reduction, nitrate => ammonia	Nitrogen metabolism	0.09273
M00532	120485	41024.7	-1.5543	16.3013	Photorespiration	Other carbohydrate metabolism	0.01589
M00533	0.81743	1.05356	0.3661	-0.0962	Homoprotocatechuate degradation, homoprotocatechuate => 2-oxohept-3-enedioate	Aromatic amino acid metabolism	0.48458
M00535	892.107	80.1581	-3.4763	8.92521	Isoleucine biosynthesis, pyruvate => 2-oxobutanoate	Branched-chain amino acid metabolism	3.44E-05
M00537	1.32249	0.05195	-4.6699	-0.5412	Xylene degradation, xylene => methylbenzoate	Aromatics degradation	0.05061
M00538	1.32249	0.05195	-4.6699	-0.5412	Toluene degradation, toluene => benzoate	Aromatics degradation	0.05061
M00540	0.03504	1.05849	4.91672	-0.871	Benzoate degradation, cyclohexanecarboxylic acid =>pimeloyl-CoA	Aromatics degradation	0.76199
M00542	0.10312	3.36307	5.02732	0.79335	EHEC/EPEC pathogenicity signature, T3SS and effectors	Pathogenicity	0.85511
M00543	0.23778	0.23778	0	-2.0723	Biphenyl degradation, biphenyl => 2-oxopent-4-enoate + benzoate	Aromatics degradation	0.93854
M00545	46.9153	13.8042	-1.765	4.92409	Trans-cinnamate degradation, trans-cinnamate => acetyl-CoA	Aromatic amino acid metabolism	0.56197
M00546	85.8497	179.045	1.06044	7.04928	Purine degradation, xanthine => urea	Purine metabolism	0.56513
M00548	0.23778	0.23778	0	-2.0723	Benzene degradation, benzene => catechol	Aromatics degradation	0.93854
M00549	2716.62	1077	-1.3348	10.8894	Nucleotide sugar biosynthesis, glucose => UDP-glucose	Sugar metabolism	0.03945
M00552	180892	61210.5	-1.5633	16.8853	D-galactonate degradation, De Ley-Doudoroff pathway, D-galactonate => glycerate-3P	Other carbohydrate metabolism	4.53E-09
M00554	96.8947	161.953	0.74108	7.01596	Nucleotide sugar biosynthesis, galactose => UDP-galactose	Sugar metabolism	0.1762
M00555	40.7664	44.345	0.12139	5.41128	Betaine biosynthesis, choline => betaine	Serine and threonine metabolism	0.09165

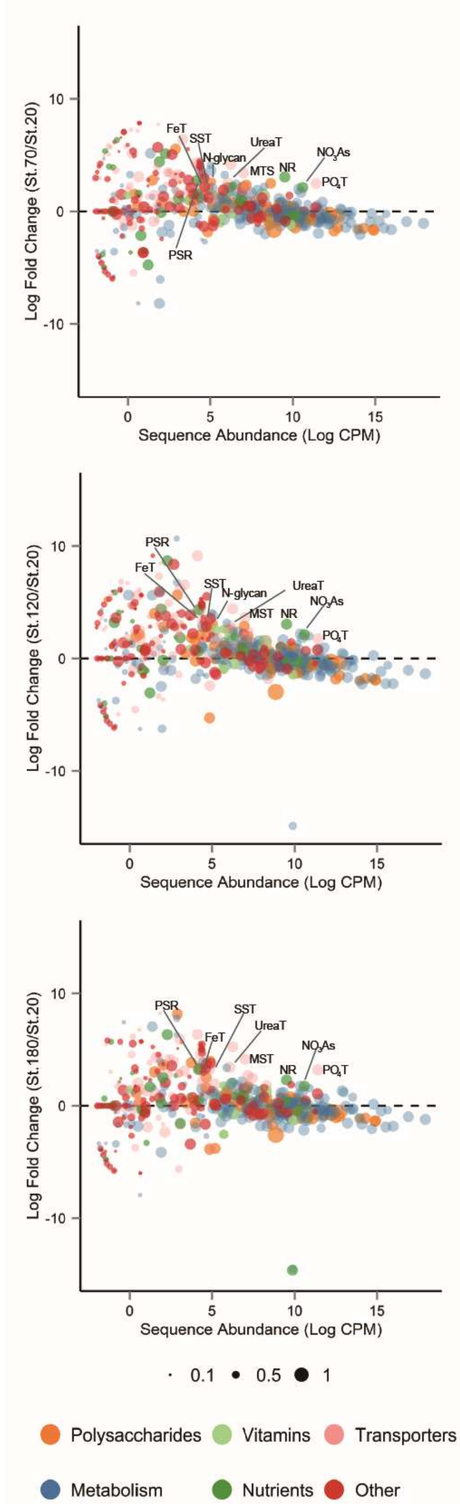
M00563	1.2607	3.04715	1.27324	1.10697	Methanogenesis, methylamine/dimethylamine/trimethylamine => methane	Methane metabolism	0.99997
M00565	2980.76	1908.5	-0.6432	11.2554	Trehalose biosynthesis, D-glucose 1P => trehalose	Sugar metabolism	0.7423
M00567	1.2607	3.04715	1.27324	1.10697	Methanogenesis, CO2 => methane	Methane metabolism	0.99997
M00568	0.33991	0.6843	1.00949	-0.9655	Catechol ortho-cleavage, catechol => 3-oxoadipate	Aromatics degradation	0.98543
M00569	46.8968	9.0555	-2.3726	4.80613	Catechol meta-cleavage, catechol => acetyl-CoA / 4-methylcatechol => propanoyl-CoA	Aromatics degradation	0.52732
M00570	5015.71	2442.75	-1.0379	11.8647	Isoleucine biosynthesis, threonine => 2-oxobutanoate => isoleucine	Branched-chain amino acid metabolism	0.1626
M00571	0.25551	51.4409	7.6534	4.69199	AlgE-type Mannuronan C-5-Epimerase transport system	Bacterial secretion system	0.00104
M00572	2817.15	1093.38	-1.3654	10.9331	Pimeloyl-ACP biosynthesis, BioC-BioH pathway, malonyl-ACP => pimeloyl-ACP	Cofactor and vitamin biosynthesis	0.011
M00573	212.804	98.8885	-1.1056	7.28398	Biotin biosynthesis, BioI pathway, long-chain-acyl-ACP => pimeloyl-ACP => biotin	Cofactor and vitamin biosynthesis	0.55826
M00575	1.11134	52.1109	5.55121	4.73396	Pertussis pathogenicity signature 2, T1SS	Pathogenicity	0.00853
M00577	212.804	98.8885	-1.1056	7.28398	Biotin biosynthesis, BioW pathway, pimelate => pimeloyl-CoA => biotin	Cofactor and vitamin biosynthesis	0.55826
M00579	3.72295	83.0529	4.47951	5.43922	Phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA => acetate	Carbon fixation	0.46956
M00580	3879.13	755.521	-2.3602	11.1782	Pentose phosphate pathway, archaea, fructose 6P => ribose 5P	Central carbohydrate metabolism	8.66E-09
M00581	0.06809	2.20465	5.01689	0.18444	Biotin transport system	Metallic cation, iron-siderophore and vitamin B12 transport system	0.50422
M00582	0.50852	3.29839	2.69739	0.92862	Energy-coupling factor transport system	Metallic cation, iron-siderophore and vitamin B12 transport system	0.92609
M00586	0.02528	2.6631	6.71872	0.42674	Putative amino-acid transport system	Phosphate and amino acid transport system	0.39972
M00595	1392.64	1071.72	-0.3779	10.267	Thiosulfate oxidation by SOX complex, thiosulfate => sulfate	Sulfur metabolism	0.45186
M00596	1704.53	1573.3	-0.1156	10.6785	Dissimilatory sulfate reduction, sulfate => H2S	Sulfur metabolism	0.78843
M00597	0.09272	1.56291	4.07514	-0.2726	Anoxygenic photosystem II	Photosynthesis	0.90932
M00599	0.65714	0.02628	-4.6441	-1.5492	Inositol-phosphate transport system	Saccharide, polyol, and lipid transport system	0.12287
M00602	2.61556	2.61556	0	1.38712	Arabinosaccharide transport system	Saccharide, polyol, and lipid transport system	1
M00605	3.00516	3.95918	0.39776	1.79999	Glucose/mannose transport system	Saccharide, polyol, and lipid transport system	1
M00606	2.61556	2.61556	0	1.38712	N,N'-Diacetylchitobiose transport system	Saccharide, polyol, and lipid transport system	1
M00607	3.14047	23.1086	2.87938	3.71419	Glycerol transport system	Saccharide, polyol, and lipid transport system	0.35951
M00608	0.06316	0.60312	3.25537	-1.5858	2-Oxocarboxylic acid chain extension, 2-oxoglutarate => 2-oxoadipate => 2-oxopimelate => 2-oxosuberate	Methane metabolism	0.88239
M00609	21307.1	13027.4	-0.7098	14.0674	Cysteine biosynthesis, methionine => cysteine	Cysteine and methionine metabolism	0.13265
M00615	639.796	2955.56	2.20775	10.8119	Nitrate assimilation	Metabolic capacity	0.06319
M00620	891.242	973.939	0.12801	9.8651	Incomplete reductive citrate cycle, acetyl-CoA => oxoglutarate	Carbon fixation	0.65518

M00622	0.71257	16.3755	4.52237	3.09491	Nicotinate degradation, nicotinate => fumarate	Cofactor and vitamin biosynthesis	0.52531
M00625	0.58089	0.03358	-4.1127	-1.7026	Methicillin resistance	Drug resistance	0.1404
M00627	1.47587	2.5483	0.78797	1.00869	beta-Lactam resistance, Bla system	Drug resistance	0.76136
M00628	0.14996	2.30936	3.94486	0.29826	beta-Lactam resistance, AmpC system	Drug resistance	0.882
M00631	8.07671	126.649	3.97093	6.07388	D-Galacturonate degradation (bacteria), D-galacturonate => pyruvate + D-glyceraldehyde 3P	Other carbohydrate metabolism	0.51732
M00632	550.512	476.042	-0.2097	9.00359	Galactose degradation, Leloir pathway, galactose => alpha-D-glucose-1P	Other carbohydrate metabolism	0.04125
M00633	73.6306	25.1821	-1.5479	5.62662	Semi-phosphorylative Entner-Doudoroff pathway, gluconate/galactonate => glycerate-3P	Central carbohydrate metabolism	0.09149
M00646	0.25551	51.4409	7.6534	4.69199	Multidrug resistance, efflux pump AcrAD-TolC	Drug efflux transporter/pump	0.00104
M00647	0.25551	51.4409	7.6534	4.69199	Multidrug resistance, efflux pump AcrAB-TolC/SmeDEF	Drug efflux transporter/pump	0.00104
M00651	0.14907	0.35642	1.25761	-1.9843	Vancomycin resistance, D-Ala-D-Lac type	Drug resistance	0.93854
M00660	0.10312	3.36307	5.02732	0.79335	Xanthomonas spp. pathogenicity signature, T3SS and effectors	Plant pathogenicity	0.85511
M00662	0.03522	0.99994	4.82742	-0.9501	Hk1-Rrp1 (glycerol uptake and utilization) two-component regulatory system	Two-component regulatory system	0.68301
M00668	2.43527	14.3833	2.56224	3.07198	Tetracycline resistance, TetA transporter	Drug efflux transporter/pump	0.43277
M00669	5.38999	6.73085	0.32051	2.59942	gamma-Hexachlorocyclohexane transport system	Saccharide, polyol, and lipid transport system	0.98182
M00670	5.38999	6.73085	0.32051	2.59942	Mce transport system	Saccharide, polyol, and lipid transport system	0.98182
M00672	3.12426	4.43008	0.50382	1.91731	Penicillin biosynthesis, aminoadipate + cycteine + valine => penicillin	Biosynthesis of secondary metabolites	0.76736
M00673	3.16808	5.87067	0.88991	2.17612	Cephamycin C biosynthesis, aminoadipate + cycteine + valine => cephamycin C	Biosynthesis of secondary metabolites	0.70887
M00676	49.6528	80.2215	0.69211	6.02097	PI3K-Akt signaling	Cell signaling	0.80158
M00677	240.648	163.361	-0.5589	7.65824	Wnt signaling	Cell signaling	0.98489
M00678	320.862	193.286	-0.7312	8.00604	Hedgehog signaling	Cell signaling	0.65391
M00679	1.12263	4.10355	1.86999	1.38576	BMP signaling	Cell signaling	0.99985
M00680	1.77943	7.8706	2.14506	2.27053	TGF-beta signaling	Cell signaling	0.80894
M00681	0.68157	8.14209	3.57846	2.14138	Activin signaling	Cell signaling	0.99295
M00682	330.126	266.093	-0.3111	8.2197	Notch signaling	Cell signaling	0.28742
M00683	98.9315	139.261	0.49329	6.89598	Hippo signaling	Cell signaling	0.76319
M00684	0.86096	12.155	3.81947	2.70221	JAK-STAT signaling	Cell signaling	0.98142
M00685	0.30594	2.61633	3.09621	0.54709	Apoptotic machinery	Cell signaling	0.86988
M00686	35.9812	23.2132	-0.6323	4.88739	Toll-like receptor signaling	Cell signaling	0.98639
M00687	167.556	233.193	0.47688	7.64656	MAPK (ERK1/2) signaling	Cell signaling	0.59631
M00688	60.3747	82.9636	0.45853	6.16328	MAPK (JNK) signaling	Cell signaling	0.96255
M00689	66.2285	81.4283	0.29808	6.2061	MAPK (p38) signaling	Cell signaling	0.98546
M00690	12.0071	38.8459	1.69387	4.66826	MAPK (ERK5) signaling	Cell signaling	0.31907
M00691	33.6389	96.7599	1.52428	6.02679	DNA damage-induced cell cycle checkpoints	Cell signaling	0.75089
M00692	602.57	385.548	-0.6442	8.94854	Cell cycle - G1/S transition	Cell signaling	0.77578
M00693	1326.36	845.64	-0.6494	10.0848	Cell cycle - G2/M transition	Cell signaling	0.70036
M00694	213.451	448.365	1.07077	8.37028	cGMP signaling	Cell signaling	0.87301

M00695	502.286	594.522	0.24322	9.0991	cAMP signaling	Cell signaling	0.61246
M00696	0.25551	51.4409	7.6534	4.69199	Multidrug resistance, efflux pump AcrEF-TolC	Drug resistance	0.00104
M00697	0.25551	51.4409	7.6534	4.69199	Multidrug resistance, efflux pump MdtEF-TolC	Drug resistance	0.00104
M00701	0.0901	0.47944	2.41176	-1.8121	Multidrug resistance, efflux pump EmrAB	Drug efflux transporter/pump	0.93854
M00709	0.45062	58.9662	7.03182	4.8928	Macrolide resistance, MacAB-TolC transporter	Drug efflux transporter/pump	0.0113
M00713	0.05594	3.63104	6.02037	0.88244	Fluoroquinolone resistance, efflux pump LfrA	Drug efflux transporter/pump	0.65113
M00714	0.05594	3.63104	6.02037	0.88244	Multidrug resistance, efflux pump QacA	Drug resistance	0.65113
M00717	0.04125	0.87559	4.40783	-1.1253	Multidrug resistance, efflux pump NorA	NA	0.51344
M00720	0.25551	51.4409	7.6534	4.69199	Multidrug resistance, efflux pump VexEF-TolC	Drug efflux transporter/pump	0.00104
M00721	0.45245	8.30811	4.1987	2.13102	Cationic antimicrobial peptide (CAMP) resistance, arnBCADTEF operon	Drug resistance	0.91951
M00723	0.19511	7.52526	5.26935	1.94867	Cationic antimicrobial peptide (CAMP) resistance, phosphoethanolamine transferase EptB	Drug resistance	0.63992
M00724	0.19511	7.52526	5.26935	1.94867	Cationic antimicrobial peptide (CAMP) resistance, palmitoyl transferase PagP	Drug resistance	0.63992
M00726	0.08228	0.50694	2.62319	-1.7631	Cationic antimicrobial peptide (CAMP) resistance, lysyl-phosphatidylglycerol (L-PG) synthase MprF	Drug resistance	0.90576
M00727	1.61967	1.32941	-0.2849	0.56026	Cationic antimicrobial peptide (CAMP) resistance, N-acetylmuramoyl-L-alanine amidase AmiA and AmiC	Drug resistance	0.99553
M00728	0.81291	18.8832	4.53786	3.29984	Cationic antimicrobial peptide (CAMP) resistance, envelope protein folding and degrading factors DegP and DsbA	Drug resistance	0.97584
M00729	32.3816	29.1144	-0.1534	4.94242	Fluoroquinolone resistance, gyrase-protecting protein Qnr	Drug resistance	0.95402
M00736	0.23778	0.23778	0	-2.0723	Nocardicin A biosynthesis, L-pHPG + arginine + serine => nocardicin A	Biosynthesis of secondary metabolites	0.93854
M00740	10654.2	7320.42	-0.5414	13.1337	Methylaspartate cycle	Other carbohydrate metabolism	0.39254
M00741	329.991	327.273	-0.0119	8.36033	Propanoyl-CoA metabolism, propanoyl-CoA => succinyl-CoA	Other carbohydrate metabolism	0.93979
M00742	1822.72	3393.96	0.89688	11.3489	Aminoglycoside resistance, protease FtsH	Drug resistance	0.68087
M00743	1.29027	1.42758	0.1459	0.44247	Aminoglycoside resistance, protease HtpX	Drug resistance	0.94388
M00744	0.19511	7.52526	5.26935	1.94867	Cationic antimicrobial peptide (CAMP) resistance, protease PgtE	Drug resistance	0.63992
M00745	0.34311	2.09679	2.61146	0.28682	Imipenem resistance, repression of porin OprD	Drug resistance	0.99806
M00760	0.51556	6.06296	3.55582	1.71776	Erythromycin resistance, macrolide 2'-phosphotransferase I MphA	Drug resistance	0.14046
M00761	0.25733	0.78286	1.60512	-0.9432	Undecaprenylphosphate alpha-L-Ara4N biosynthesis, UDP-GlcA => Undecaprenyl phosphate alpha-L-Ara4N	NA	0.98544
M00763	1.31453	0.86361	-0.6061	0.1231	Ornithine biosynthesis, mediated by LysW, glutamate => ornithine	Arginine and proline metabolism	0.77096

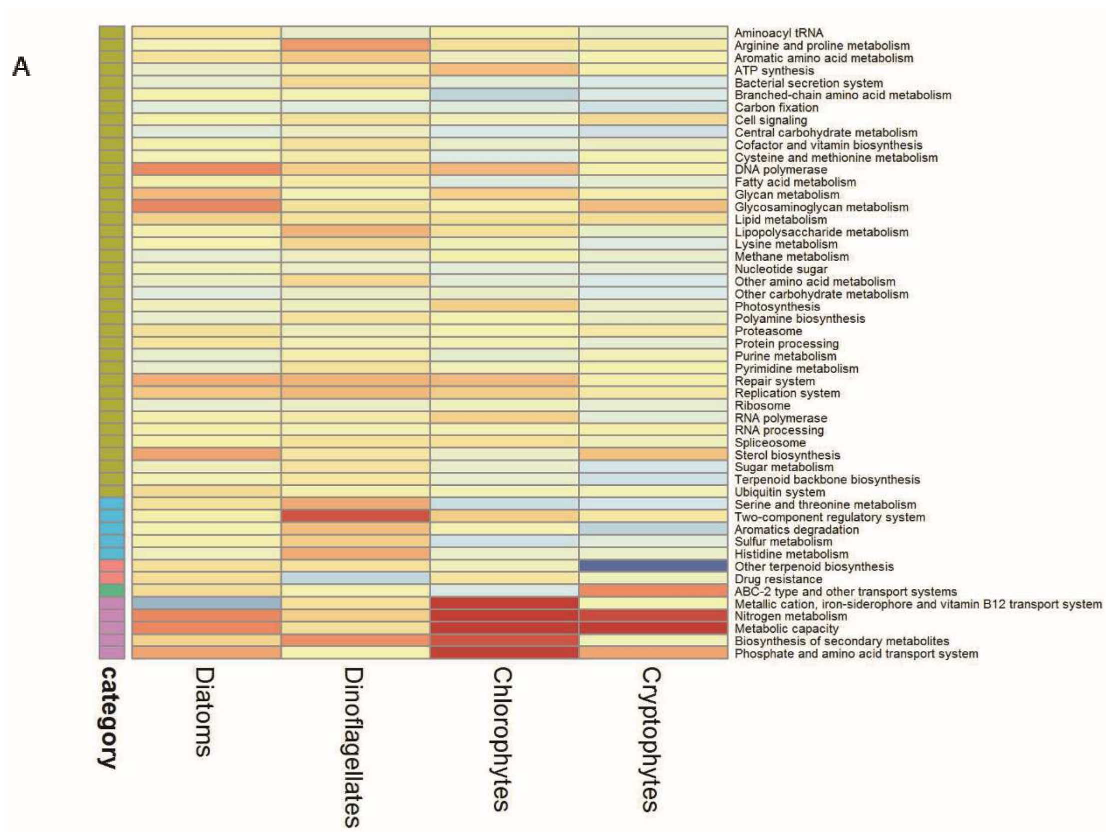
APPENDIX 4: MA PLOT BETWEEN UPPER AND LOWER STATIONS

Differential expression analysis between the upper and lower estuary stations for (A) station 20 versus station 70, (B) station 20 versus station 120 and (C) station 20 versus station 180.

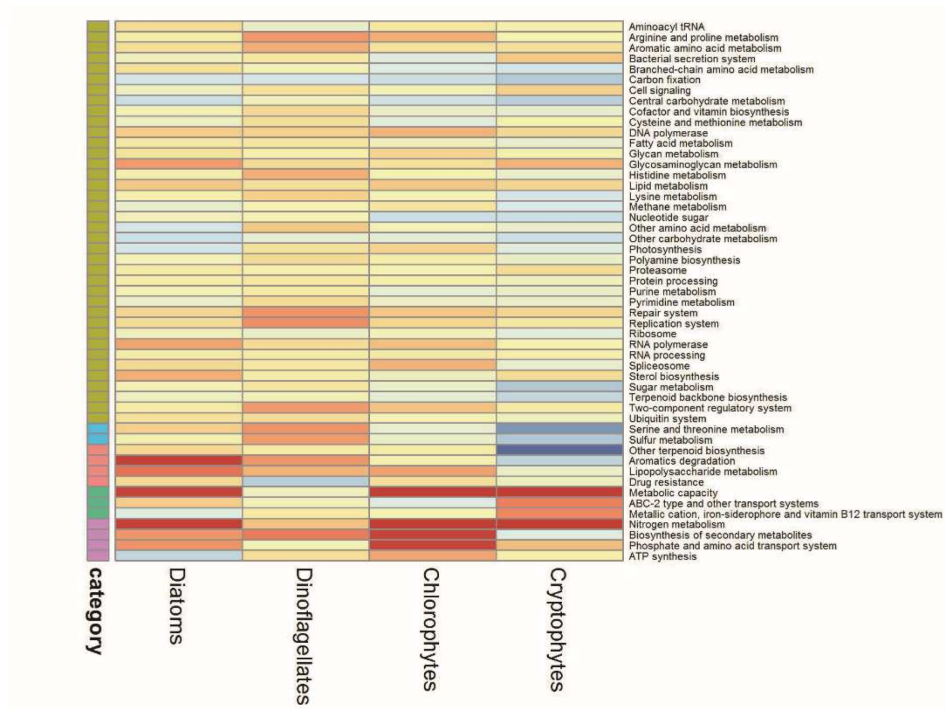


APPENDIX 5: HEATMAP FOR EXPRESSION PROFILES

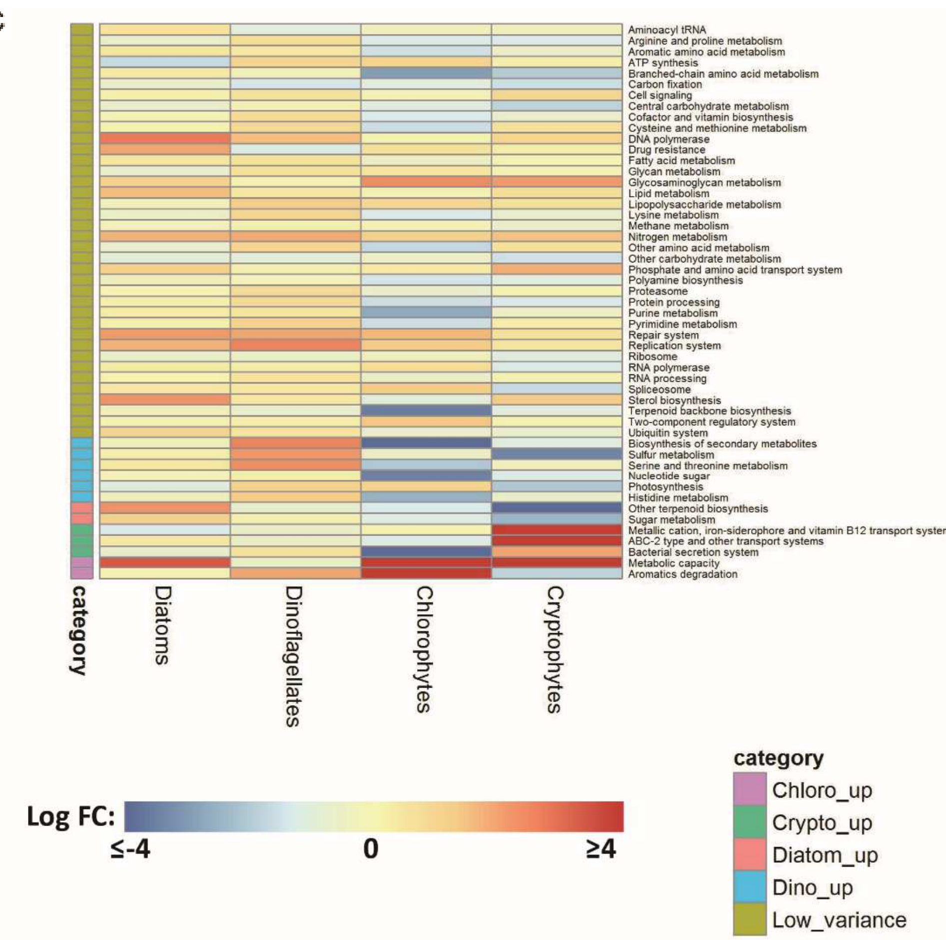
Heatmap for the expression profiles based on KEGG clas3 between a) station 20 and station 70, b) station 20 and station 120, c) station 20 and station 180. Each row represents the expression level of a KEGG clas3 with warmer color (positive fold change) indicating over-representation in the lower estuary stations (station 70, 120, 180) and cooler color (negative fold change) indicating over-representation at station 20.



B


Aminoacyl tRNA
Arginine and proline metabolism
Aromatic amino acid metabolism
Bacterial secretion system
Branched-chain amino acid metabolism
Carbon fixation
Cell signaling
Central carbohydrate metabolism
Cofactor and vitamin biosynthesis
Cysteine and methionine metabolism
DNA polymerase
Fatty acid metabolism
Glycan metabolism
Glycosaminoglycan metabolism
Histidine metabolism
Lipid metabolism
Lysine metabolism
Methane metabolism
Nucleotide sugar
Other amino acid metabolism
Other carbohydrate metabolism
Photosynthesis
Polyamine biosynthesis
Proteasome
Protein processing
Purine metabolism
Pyrimidine metabolism
Repair system
Replication system
Ribosome
RNA polymerase
RNA processing
Spliceosome
Sterol biosynthesis
Sugar metabolism
Terpenoid backbone biosynthesis
Two-component regulatory system
Ubiquitin system
Serine and threonine metabolism
Sulfur metabolism
Other terpenoid biosynthesis
Aromatics degradation
Lipopolysaccharide metabolism
Drug resistance
Metabolic capacity
ABC-2 type and other transport systems
Metallic cation, iron-siderophore and vitamin B12 transport system
Nitrogen metabolism
Biosynthesis of secondary metabolites
Phosphate and amino acid transport system
ATP synthesis
category
Diatoms
Dinoflagellates
Chlorophytes
Cryptophytes

C


Aminoacyl tRNA
Arginine and proline metabolism
Aromatic amino acid metabolism
ATP synthesis
Branched-chain amino acid metabolism
Carbon fixation
Cell signaling
Central carbohydrate metabolism
Cofactor and vitamin biosynthesis
Cysteine and methionine metabolism
DNA polymerase
Drug resistance
Fatty acid metabolism
Glycan metabolism
Glycosaminoglycan metabolism
Lipid metabolism
Lipopolysaccharide metabolism
Lysine metabolism
Methane metabolism
Nitrogen metabolism
Other amino acid metabolism
Other carbohydrate metabolism
Phosphate and amino acid transport system
Polyamine biosynthesis
Proteasome
Protein processing
Purine metabolism
Pyrimidine metabolism
Repair system
Replication system
Ribosome
RNA polymerase
RNA processing
Spliceosome
Sterol biosynthesis
Terpenoid backbone biosynthesis
Two-component regulatory system
Ubiquitin system
Biosynthesis of secondary metabolites
Sulfur metabolism
Serine and threonine metabolism
Nucleotide sugar
Photosynthesis
Histidine metabolism
Other terpenoid biosynthesis
Sugar metabolism
Metallic cation, iron-siderophore and vitamin B12 transport system
ABC-2 type and other transport systems
Bacterial secretion system
Metabolic capacity
Aromatics degradation
category
Diatoms
Dinoflagellates
Chlorophytes
Cryptophytes
Log FC:
≤-4
0
≥4
category
Chloro_up
Crypto_up
Diatom_up
Dino_up
Low_variance

REFERENCES

Anderson, M.J. & Walsh, D.C.I., 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecological Monographs*, 83(4), pp.557–574. Available at: http://doi.wiley.com/10.1890/12-2010.1 [Accessed November 1, 2016].

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Azam, F. et al., 1983. The Ecological Role of Water-Column Microbes in the Sea . *Marine Ecology Progress Series*, 10, pp.257–263.

Brussaard, C.P.D., 2003. Viral control of phytoplankton populations--a review. *The Journal of eukaryotic microbiology*, 51(2), pp.125–38.

Carstensen, J., Henriksen, P. & Heiskanen, A.-S., 2007. Summer algal blooms in shallow estuaries: Definition, mechanisms, and link to eutrophication. *Limnology and Oceanography*, 52(1), pp.370–384.

Cira, E.K. et al., 2016. Effects of Nitrogen Availability and Form on Phytoplankton Growth in a Eutrophied Estuary (Neuse River Estuary, NC, USA) C. J. Gobler, ed. *PLOS ONE*, 11(8), p.e0160663.

Clarke, K.R. & Warwick, R.M., 2001. Change in marine communities. An approach to statistical analysis and interpretation. *Natural Environment Research Council*, pp.1–172.

Cloern, J., 2001. Our evolving conceptual model of the coastal eutrophication problem. *Marine Ecology Progress Series*, 210, pp.223–253.

Cooper, E.D. et al., 2014. Metatranscriptome profiling of a harmful algal bloom. *Harmful Algae*, 37, pp.75–83.

Cooper, M.B. & Smith, A.G., 2015. Exploring mutualistic interactions between microalgae and bacteria in the omics age. *Current Opinion in Plant Biology*, 26, pp.147–153.

Croft, M.T. et al., 2005. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature*, 438(November), pp.90–93.

Day, J.W. et al., 2012. *Estuarine Ecology* 2nd ed., New York, NY: John Wiley.

Delbarre-Ladrat, C. et al., 2014. Exopolysaccharides produced by marine bacteria and their applications as glycosaminoglycan-like molecules. *Frontiers in Chemistry*, 2(October), pp.1–15.

Demir, E. et al., 2008. Assessment of Microzooplankton Grazing on Heterosigma akashiwo Using a Species- Specific Approach Combining Quantitative Real-Time PCR (QPCR) and Dilution Methods. *Microbial Ecology*, 55(4), pp.583–594.

Fawcett, S.E. & Ward, B.B., Phytoplankton succession and nitrogen utilization during the development of an upwelling bloom. *Marine Ecology Progress Series*, 428, pp.13–31.

Gifford, S.M. et al., 2012. Expression patterns reveal niche diversification in a marine microbial assemblage. *The ISME Journal*, 7(2), pp.281–298.

Giordano, M., Beardall, J. & Raven, J. a, 2005. CO2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annual review of plant biology*, 56, pp.99–131.

Graham, S. & Strom, S., 2010. Growth and grazing of microzooplankton in response to the harmful alga Heterosigma akashiwo in prey mixtures. *Aquatic Microbial Ecology*, 59(2), pp.111–124.

Hall, N.S. et al., 2012. Effects of climatic variability on phytoplankton community structure and bloom development in the eutrophic , microtidal , New River Estuary , North Carolina , USA. *Estuarine, Coastal and Shelf Science*, 117, pp.70–82.

Hall, N.S. & Pearl, H., 2011. Vertical migration patterns of phytoflagellates in relation to light and nutrient availability in a shallow microtidal estuary. *Marine Ecology Progress Series*, 425, pp.1–19.

Hatch, M.D. & Slack, C.R., 1968. A new enzyme for the interconversion of pyruvate and phosphopyruvate and its role in the C4 dicarboxylic acid pathway of photosynthesis. *Biochem J*, 106(1), pp.141–146.

Howarth, R.W., Billen, G. & Swaney, D., 1997. Regional nitrogen budgets and riverine N and P fluxes for the drainages to the North Atlantic Ocean: natural and human influences. *Oceanographic Literature Review*, 5(44), p.448.

Jones, P.M. & George, A.M., 2004. The ABC transporter structure and mechanism: perspectives on recent research. *Cellular and molecular life sciences : CMLS*, 61(6), pp.682–99.

Kang, L.-K. et al., 2007. Influences of nitrogen deficiency on the transcript levels of ammonium transporter , nitrate trans ... *Phycologia*, pp.521–533.

Keeling, P.J. 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.

Kennish, M.J. & Paerl, H.W., 2010. *Coastal lagoons : critical habitats of environmental change*, Taylor & Francis.

Langfelder, P. & Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9(1), p.559.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), pp.2078–9.

Liu, Z. et al., 2015. Changes in gene expression of Prymnesium parvum induced by nitrogen and phosphorus limitation. *Frontiers in microbiology*, 6, p.631.

Maheswari, U. et al., 2010. Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome biology*, 11(8), p.R85.

Mallin, M.A. & Paerl, H.W., 1992. Effects of variable irradiance on phytoplankton productivity in shallow estuaries. *Limnology and Oceanography*, 37(1), pp.54–62. A

Marchetti, A. et al., 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proceedings of the National Academy of Sciences*, 109(6), pp.E317–E325.

Nielsen, R. et al., 2005. A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. *PLoS Biology*, 3(6), p.e170.

Nixon, S.W., 1995. Coastal marine eutrophication: A definition, social causes, and future concerns. *Ophelia*, 41(1), pp.199–219.

Paerl, H.W., 2006. Assessing and managing nutrient-enhanced eutrophication in estuarine and coastal waters: Interactive effects of human and climatic perturbations. *Ecological Engineering*, 26, pp.40–54.

Paerl, H.W. et al., 1998. Ecosystem responses to internal and watershed organic matter loading: consequences for hypoxia in the eutrophying Neuse River Estuary, North Carolina, USA. *Marine Ecology Progress Series*, 166, pp.17–25.

Paerl, H.W., Hall, N.S., Peierls, B.L. & Rossignol, K.L., 2014. Evolving Paradigms and Challenges in Estuarine and Coastal Eutrophication Dynamics in a Culturally and Climatically Stressed World. *Estuaries and Coasts*, 37(2), pp.243–258.

Paerl, H.W., Hall, N.S., Peierls, B.L., Rossignol, K.L., et al., 2014. Hydrologic Variability and Its Control of Phytoplankton Community Structure and Function in Two Shallow, Coastal, Lagoonal Ecosystems: The Neuse and New River Estuaries, North Carolina, USA. *Estuaries and Coasts*, 37(S1), pp.31–45.

Paerl, H.W. et al., 1995. Nitrogen loading sources and eutrophication of the Neuse River Estuary, North Carolina: Direct and indirect roles of atmospheric deposition. *Rep*, 291.

Paerl, H.W. et al., 2010. Phytoplankton community indicators of short- and long-term ecological change in the anthropogenically and climatically impacted neuse river estuary, North Carolina, USA. *Estuaries and Coasts*, 33, pp.485–497.

Peierls, B.L., Hall, N.S. & Paerl, H.W., 2012. Non-monotonic Responses of Phytoplankton Biomass Accumulation to Hydrologic Variability: A Comparison of Two Coastal Plain North Carolina Estuaries. *Estuaries and Coasts*, 35(6), pp.1376–1392.

Piehler, M.F. et al., 2004. Impacts of inorganic nutrient enrichment on phytoplankton community structure and function in Pamlico Sound, NC, USA. *Estuarine, Coastal and Shelf Science*, 61, pp.197–209.

Pinckney, J.L. et al., 1998. Annual cycles of phytoplankton community-structure and bloom dynamics in the Neuse River Estuary, North Carolina. *Marine Biology*, 131(2), pp.371–381.

Pinckney, J.L. et al., 1997. Environmental controls of phytoplankton bloom dynamics in the Neuse River Estuary, North Carolina, U.S.A. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(11), pp.2491–2501.

Pinckney, J.L. et al., 1996. Flow scintillation counting of 14 C-labeled microalgal photosynthetic pigments. *Journal of Plankton Research*, 18(10), pp.1867–1880.

Reed, R.E. et al., 2004. Seasonal physical–chemical structure and acoustic Doppler current profiler flow patterns over multiple years in a shallow, stratified estuary, with implications for lateral variability. *Estuarine, Coastal and Shelf Science*, 60(4), pp.549–566.

Reinfelder, J.R., Kraepiel, A.M. & Morel, F.M., 2000. Unicellular C4 photosynthesis in a marine diatom. *Nature*, 407(6807), pp.996–9.

Rinta-Kanto, J.M. et al., 2012. Bacterial community transcription patterns during a marine phytoplankton bloom. *Environmental Microbiology*, 14, pp.228–239.

Robertson, G. et al., 2010. De novo assembly and analysis of RNA-seq data. *Nature methods*, 7(11), pp.909–12. Available at: http://dx.doi.org/10.1038/nmeth.1517 [Accessed August 7, 2015].

Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139–140.

Robinson, M.D. & Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, 11, p.R25.

Schruth, D., 2013. caroline: A Collection of Database, Data Structure, Visualization, and Utility Functions for R. Available at: http://cran.r-project.org/package=caroline.

Seymour, J.R. et al., 2017. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nature Microbiology*, 2(May), p.17065.

Sharp, J.H. et al., 1984. Estuarine Interaction of Nutrients, Organics, and Metals: A Case Study in the Delaware Estuary. *The Estuary as a Filter, Academic Press, Orlando FL. 1984. P 241-258, 10 fig, 23 ref. NA 80 AA-D-00106.*, pp.241–258.

Simpson, J.T. et al., 2009. ABySS: a parallel assembler for short read sequence data. *Genome research*, 19(6), pp.1117–23.

Smayda, T.J., 1997. Harmful algal blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnology and Oceanography*, 42(5_part_2), pp.1137–1153.

Smith, G.J., Zimmerman, R.C. & Alberte, R.S., 1992. Molecular and physiological responses of diatoms to variable levels of irradiance and nitrogen availability: Growth of Skeletonema costatum in simulated upwelling conditions. *Limnology and Oceanography*, 37(5), pp.989–1007.

Song, B. & Ward, B.B., 2007. MOLECULAR CLONING AND CHARACTERIZATION OF HIGH-AFFINITY NITRATE TRANSPORTERS IN MARINE PHYTOPLANKTON. *Journal of Phycology*, 43(3), pp.542–552.

Twomey, L.J., Piehler, M.F. & Paerl, H.W., 2005. Phytoplankton uptake of ammonium, nitrate and urea in the Neuse River Estuary, NC, USA. *Hydrobiologia*, 533(1–3), pp.123–134.

Valdes-Weaver, L.M. et al., 2006. Long-term temporal and spatial trends in phytoplankton biomass and class-level taxonomic composition in the hydrologically variable Neuse-Pamlico estuarine continuum, North Carolina, U.S.A. *Limnology and Oceanography*, 51(3), pp.1410–1420.

Walz, N. & Welker, M., 1998. Plankton Development in a Rapidly Flushed Lake in the River Spree System (Neuendorfer See, Northeast Germany). *J Plankton Res*, Vol 20(Iss 11), pp.2071–2087.

Wear, E.K. et al., 2015. Roles of diatom nutrient stress and species identity in determining the short- and long-term bioavailability of diatom exudates to bacterioplankton. *Marine Chemistry*, 177, pp.335–348.

Wei, Y. et al., 2001. High-density microarray-mediated gene expression profiling of Escherichia coli. *Journal of bacteriology*, 183(2), pp.545–56.

Welschmeyer, N.A., 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. *Limnology and Oceanography*, 39(8), pp.1985–1992.