

Microbiome Community Dynamics in Large Outdoor Algae Raceway Ponds

by

Courtney Jean Swink

Marine Science and Conservation
Duke University

Date: _____

Approved:

Zackary I Johnson, Advisor

Dana E Hunt

Lawrence A David

Emiley A Eloë-Fadrosh

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science
in Marine Science and Conservation
in the Graduate School
of Duke University

2020

ABSTRACT

Microbiome Community Dynamics in Large Outdoor Algae Raceway Ponds

by

Courtney Jean Swink

Marine Science and Conservation
Duke University

Date: _____

Approved:

Zackary I Johnson, Advisor

Dana E Hunt

Lawrence A David

Emiley A Eloë-Fadrosh

An abstract of a thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science
in Marine Science and Conservation
in the Graduate School
of Duke University

2020

Copyright by
Courtney Jean Swink
2020

Abstract

Marine microalgae are photosynthetic microbes that are a potential source of fuels, animal feed, and other specialized products. Large scale cultivation of microalgae occurs in open, outdoor raceway ponds, which are exposed to the natural environment and these cultures quickly become a complex milieu of microbes. Microalgae interact with attached and free-floating bacteria found in their medium, with both positive and negative outcomes. To investigate the diversity and dynamics of microbes associated with these systems, samples were collected during multiple growth cycles of two biofuel-relevant microalgae strains, *Desmodesmus* sp. and *Oocystis* sp. in ~4,500 L outdoor raceway ponds. Microbiome community composition and diversity was dramatically different between ponds from the two algae and from the natural microbiome of the treated seawater used in pond medium. In spite of variable environments, the pond microbiomes were most similar to their inoculum PBR (photobioreactor) communities suggesting the importance of priority effects or environmental conditioning by the host algae. Ponds when both algae strains were grown were dominated by *Rhodobacteraceae* and *Saprospiraceae* while unhealthy microbiomes were dominated by *Cytophagaceae* and *Puniceicoccaceae*. Microbiome change was variable over time and resulted in different community structures at the time of algae harvest. Variation in the microbiome community structure was driven by the strain of algae grown, time, pond temperature

and percent oxygen saturation. These results provide insight into this industrial ecology and are a foundation for future microbiome research to improve microalgae production.

Contents

Abstract	iv
List of Tables	vii
List of Figures	viii
1. Introduction	1
1.1 Study Objectives	5
2. Methods.....	7
2.1 Study System.....	7
2.2 Sample Collection.....	9
2.3 DNA Extraction and Library Preparation.....	10
2.4 Sequence Processing	11
2.5 Data Analysis	12
3. Results.....	15
3.1 Microalgae Growth in the Cultivation System Environment	15
3.2 Microbiome Variation Across Host Strains of Algae	18
3.3 Microbiome Variation Over Time	27
3.4 Environmental Drivers of Microbiome Change.....	33
4. Discussion	36
5. Conclusion	48
References	52

List of Tables

Table 1: Variance partitioning analysis on pond environmental drivers.....	35
--	----

List of Figures

Figure 1: MAGIC cultivation system.....	8
Figure 2: Rarefaction curves.	12
Figure 3: Cultivation environment.	17
Figure 4: Algae specific growth rate for each pond run.	18
Figure 5: Ordination based on Bray Curtis Dissimilarity.....	19
Figure 6: Differences in Shannon diversity across strain and container type.	21
Figure 7: Overlap of the top 150 ASVs between strains.	23
Figure 8: Microbiome variation between algal strains summed at the family level.	24
Figure 9: Abundance heatmap for top 150 taxa across all samples.	26
Figure 10: Microbiome variation over time in each production pond during the growth of both algal strains.....	29
Figure 11: Changes in Shannon diversity over time in production ponds.	30
Figure 12: Abundance change for top 150 ASVs during each pond run.....	32
Figure 13: Trajectory for <i>Phaeodactylibacter sp.</i> during each pond run.....	33
Figure 14: Environmental drivers of microbiome change in production ponds.....	34

1. Introduction

Marine microalgae (phytoplankton) are single celled, photosynthetic organisms that account for roughly half of all global primary production (Behrenfeld et al. 2001). In both natural and laboratory environments, microalgae do not typically grow axenically, but are part of a broader microbial community. In algae cultures, bacteria attached to algae cell surfaces vary significantly even between closely related microalgae lineages (Baker and Kemp 2014). For example, the relationship between diatoms and their associated attached bacteria is not stochastic (Baker and Kemp 2014) and are limited to only a few genera (Amin et al. 2012). Diatoms can even “cultivate” bacteria by releasing organic-rich compounds for their surrounding bacteria. Interactions between microalgae and surrounding bacteria can be synergistic, competitive, or parasitic (Amin et al. 2012). These interactions are found in both the natural environment and laboratory cultures and can dramatically affect microalgae growth.

While microalgae and their microbiomes may interact in many ways, it is generally thought algae benefit from the presence of bacteria in co-culture due to their synergistic interactions via symbiotic relationships (Jutson et al. 2016). For example, bacteria can recycle extracellular algae products into micronutrients (Bell and Mitchell 1972). Bacteria can be the source of some micronutrients such as vitamin B₁₂ (Kazamia et al. 2012) or fix nitrogen into a bioavailable form (Foster et al. 2011). In exchange for a

supply of dissolved organic carbon (DOC) certain types of *Gammaproteobacteria* produce compounds that have a high affinity for soluble iron, another important micronutrient for microalgae (Amin et al. 2012). Some bacterial phylogenetic groups have specificity to DOC produced by different microalgae strains (Sarmiento and Gasol 2012) and differences in the type of algae-derived DOC can lead to different functional responses in the microbiome (Beier et al. 2015). Close associations between certain types of bacteria and algae have been observed and some bacteria even release antibacterial compounds to prevent competition from other taxa (Baker and Kemp 2014), thus altering the broader microbiome. Alternatively, some microalgae are dependent on co-occurring microbes to perform certain lost metabolic functions. For example, the marine microalgae *Prochlorococcus* can survive without an oxidative stress response gene because its surrounding bacteria reduce the accumulation of hydrogen peroxide (Morris et al. 2012).

Other microbes can have negative effects on microalgae via parasitic or competitive interactions. For example, *Phaeobacter gallaeciensis*, a type of bacteria which releases selective algaecides in the presence of aged algae cells, becomes a pathogen of its algal host (Seyedsayamdost et al. 2011). Other bacteria associated with red tide causing algae (e.g. *Alteromonas* sp. and *Thalassobius aestuarii*) induce cell lysis by releasing algicidal compounds (Wang et al. 2010). Some predatory bacteria, such as *Bdellovibrio*

chlorellavorus, attack the surface of the algae cell often resulting in color change of *Chlorella* from a healthy green to yellow-green prior to death (Coder and Starr 1978). And more generally, bacteria, which have greater surface area to volume ratio, can also impede microalgae growth through competitive interactions by consuming nutrients more efficiently than algae (Amin et al 2012).

Microbial community dynamics are driven by two major categories of ecological processes: deterministic and stochastic. Deterministic processes are ecological mechanisms that change the taxa of a community in a nonrandom, predictable way often based on the organisms' ecological niches (Chase and Myers 2011), and/or priority effects (Fukami 2015). Host selection of bacterial communities in biofuel-relevant microalgae strains (Kimbrel et al. 2019) and selection based on environmental parameters (Lozupone and Knight 2007) are examples of how deterministic processes impact microbiome community structure. Priority effects occur when the order and timing of species introduction to a system create lasting effects in the future structure and function of that community (Chase 2003, Fukami 2015). Studies focused on priority effects in microbial communities have found that inoculation history is important for determining freshwater bacteria community structure (Rummens et al. 2018); early invasions by non-local bacteria are more successful than invasions in well-established communities due to priority effects (Rivett et al. 2018). Conversely, stochastic processes

are ecological mechanisms that randomly alter community structure in a non-deterministic way (Chase and Myers 2011). Examples of stochastic processes include drift, random deaths, diversification, and dispersal (particularly for non-motile organisms such as bacteria). Some studies have found that stochastic processes play important roles in community assembly as well, such as in a groundwater microbial community responding dynamically to nutrient flux (Zhou et al. 2014). Both stochastic and deterministic processes have been shown to be important in driving changes in microbial community structure in a variety of systems (Zhou and Ning 2017).

Marine microalgae are an exciting candidate for the sustainable production of biofuel, feed and other natural products because of their fast growth rates and ability to be grown in saltwater on non-arable land. Although algae have been shown to grow well using current mass cultivation techniques (i.e. raceway ponds), there remain substantial barriers to broader commercialization including high water usage, low productivity, inefficient harvesting methods (Bhujade et al. 2017) and contamination of invading weeds or predators (White and Ryan 2015). The microbiomes of microalgae can dramatically influence their productivity and can aid in crop protection (Lindemann et al. 2016). For example, a recent study demonstrated addition of nitrogen reducing bacteria to a photobioelectrical system improved system output at higher nitrogen loading concentrations (Luo et al. 2020). Another recent study found that the

microbiome of the algae *Microchloropsis salina* assisted in protection against algal predation by a marine rotifer (Fisher et al. 2019).

Synthetic bacterial communities can lead to predictable responses in plant hosts, such as changes in plant phosphate content (Herrera Paredes et al. 2018), which shows promise for microbiome science to improve biotechnology and agricultural practices. There is interest in manipulating the microbiome of algal cultures in order to improve productivity or stability of raceway pond systems, and researchers have argued for increasing efforts to engineer microbiome communities rather than focus on promoting individual microbes (Lindemann et al. 2016). Thus, similar to natural systems, it is important to understand the dynamics of microbiome change in algae-bacteria communities so that these communities could be optimized for large-scale algae cultivation.

1.1 Study Objectives

The aim of this thesis is to describe the dynamics and environmental drivers that shape microbial community structure in an industrial-relevant, outdoor marine microalgae raceway pond system. Specifically, the goals of this research are to (1) compare the raceway pond microbiomes to the microbiomes of the input sources used to inoculate them, (2) identify the dominant bacterial taxa associated with two biofuel-relevant microalgae strains, (3) describe how the pond microbiomes change with time,

and (4) identify the environmental parameters that drive variation in pond community structure.

2. Methods

2.1 Study System

As part of a larger research project conducted by the Marine AlGae Industrialization Consortium (MAGIC), the Duke University Marine Lab (Beaufort, NC) constructed five outdoor raceway ponds to cultivate marine microalgae in seawater for biofuel and other natural products. Several different marine microalgae strains were tested in these outdoor ponds over the course of two and a half years. The cultivation strategy for the larger research project was a hybrid system of covered semi-continuous photobioreactors (PBRs, Figure 1A) and larger open raceway ponds (Huntley et al. 2015, Huntley and Redalje US Patent #7,770,322). Microalgae cultures were scaled up in laboratory incubators to 20 L carboys, which were used to inoculate ~1000 L outdoor covered PBRs (usually every 1-2 weeks), which were operated in semi-continuous mode with regular dilutions to match their growth rate. These PBRs were used to inoculate two ~4500 L outdoor raceway ponds (production ponds, Figure 1A) for 2-4 days of batch growth. Algal biomass was harvested from the production ponds via membrane filtration and centrifugation for further processing and evaluation as feed and fuel products.

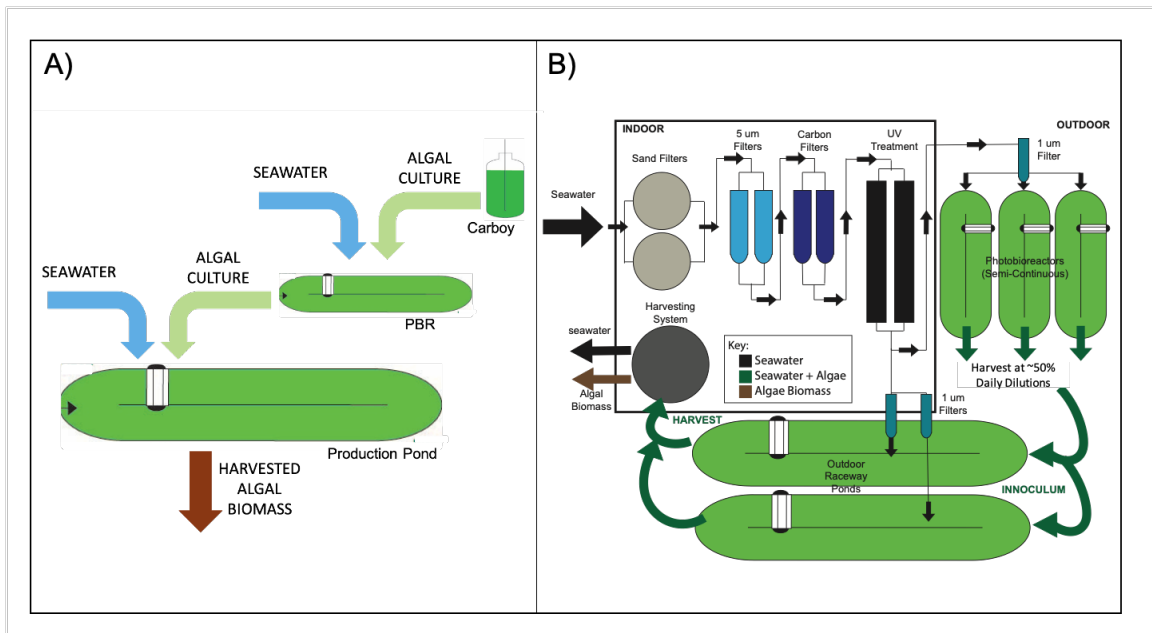


Figure 1: MAGIC cultivation system. A) Algae culture scale-up process from 20 L carboys, which seed ~1,000 L outdoor/covered raceway photobioreactors (PBRs). PBRs are used to seed ~4,500 L outdoor production ponds. Biomass is grown in production ponds for ~2-4 days before being harvested and the ponds re-inoculated. B) Flow of seawater through the MAGIC cultivation system. Coastal seawater is treated indoors through a sand filter, 5 µm filter, activated carbon filter, UV treatment and a final 1 µm point-of-use filter prior to being pumped into all outdoor PBRs and ponds. Seawater is separated from algae biomass in the harvesting system before being released into the local sewer system.

To remove predators and other microalgae prior to use, coastal seawater passed through a treatment system comprised of a sand filter, 5 µm filter, activated carbon filter, UV treatment and a final 1µm point-of-use filter prior to being pumped into both the PBRs and production ponds (Figure 1B). This water treatment process dramatically alters the microbiome of the incoming seawater, but it does not render it sterile; and much of the bacterial community remains post-filtration.

The media used to grow the algal cultures for this project used modified f media nutrient concentrations for all stages of culture scale up: 1764 μM NO_3 , 110 μM PO_4 , 1x f vitamins, 1/20x (f/2) trace metals (Andersen et al. 2005). A weather station mounted next to the raceway ponds measured photosynthetic active radiation (PAR), rainfall, wind speed and direction, and air temperature. All of the outdoor PBRs and production ponds were outfitted with the Apex Neptune system, which automatically recorded temperature, pH, and controlled CO_2 inputs. Daily sampling was conducted to quantify algal growth and productivity in all raceway ponds by the larger cultivation team. Data collected included pond temperature and dissolved oxygen saturation (ProODO probe), optical density (440 nm), turbidity (NTU), dry weight and ash free dry weight, extracted chlorophyll (100% methanol, Ritchie 2006), FIRE fluorescence (Johnson 2004), and cell counts based on flow cytometry (Johnson et al. 2010).

2.2 Sample Collection

This research used the existing infrastructure of the larger MAGIC project and added additional sampling and analyses to answer questions regarding the community dynamics of the microbiome in these raceway ponds. Samples were collected from outdoor production ponds and input sources (seawater and PBRs, Figure 1A) when a pond was restarted and bacterial 16S rDNA community analysis performed. Samples were collected daily from each of the two production raceway ponds for a minimum of 20 “pond runs” (i.e. the length of time the pond culture was grown before harvesting

and re-seeding from PBRs). Production ponds were grown for 2-4 days each, resulting in multiple samples per pond run. Samples were collected during the growth of two different microalgae strains, *Desmodesmus sp.* (C046) and *Oocystis sp.* (S002) which are both green chlorophytes (green algae) and biofuel-relevant strains (Huntley et al. 2015). Carboy, PBR and pond samples were collected by filtering ~10-20 mL of culture onto a 25 mm 0.2 μm polycarbonate filter (2 replicates each). Seawater samples were collected as the raceway ponds were being refilled, and ~2 L of seawater was filtered through a 45 mm 0.2 μm filter (Pall Supor 200). All filters were placed in 1.5 mL screw cap tubes and stored frozen at -80°C until further analysis.

2.3 DNA Extraction and Library Preparation

Genomic DNA was extracted using the entire filter for 25 mm diameter filters and from half of the 45 mm diameter filters (cut with sterile razor blades). Gentra Puregene Yeast/Bacteria Kit (QIAGEN) was used for DNA extraction from all filters following the manufacturer's instructions with additional bead beating cycles of 60 seconds (3x). Samples were cleaned with the Zymo *OneStep* PCR inhibitor removal kit. Extracted DNA samples were quantified using a Nanodrop ND-100 and diluted to ~20 ng/ μL . PCR was conducted using dual-index 16S rRNA primers 515F and 926R (V4-V5) (Parada et al. 2016). PCR reactions contained 1x EconoTaq buffer, 200 μM dNTPs, 0.5 μM forward primer, 0.5 μM reverse primer, 0.4 U EconoTaq, and 20 ng of sample DNA. The thermocycler for PCR was programmed as follows: 95°C for 30 seconds, 32 cycles of

95°C for 10 seconds, 50°C for 30 seconds, 72°C for 30 seconds, 72°C for 2 minutes.

Triplicate 20 µL reactions were pooled and run on a 1% agarose gel for purification/extraction. Libraries were pooled and diluted using a Qubit 3.0. 16S rRNA gene libraries were sequenced using 2 x 250 bp sequencing on the Illumina MiSeq at the Duke Center for Genomic and Computational Biology.

2.4 Sequence Processing

Upon receiving the raw sequencing files, barcodes were trimmed using MacQIIME v1.9.1. Files were then demultiplexed by the i5 index into fastq files for each sample using Sabre (<https://github.com/najoshi/sabre>). Amplicon sequence variants (ASVs) were determined using DADA2 version 1.12.1 (Callahan et al. 2016). Primers were trimmed, reads were quality filtered (using parameters maxN = 0, maxEE = 2, truncQ = 2), and denoised using the DADA2 error model. Reads were merged and potential chimeras removed. Taxonomy assignments for representative ASVs were made using the RDP classifier (Wang et al. 2007). 10,393 ASVs were determined from 22,212,256 reads (75% of total reads analyzed). Sequences that were assigned to chloroplast or mitochondria were removed. In order to account for variations in sequencing depth and reduce number of samples lost, samples were rarefied to 1,650 counts following Larkin et al. 2016 (Figure 2). After these steps, data from 249 microbiome samples remained for downstream analyses.

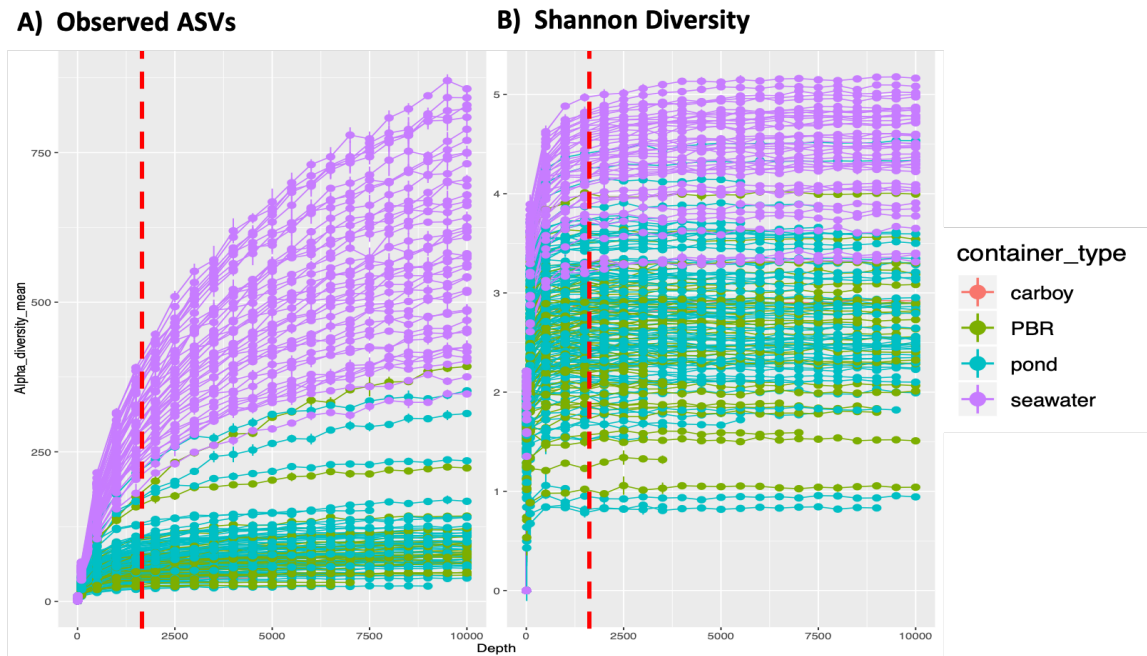


Figure 2: Rarefaction curves. Changes in alpha diversity over sequencing depth are shown for both A) observed number of ASVs and B) Shannon diversity where color of the curve corresponds to the type of container the sample was collected from. Vertical dashed red line in both plots indicates selected rarefaction depth of 1,650.

2.5 Data Analysis

Algae specific growth rate (μ) was calculated for each pond run using the equation $\mu = \ln(N_2/N_1)/(t_2 - t_1)$, where N_1 and N_2 are the biomass measured by optical density (440 nm) at pond inoculation (t_1) and harvest (t_2). Shannon diversity and Bray Curtis dissimilarity were calculated for all community samples using the R phyloseq package (version 1.28.0, McMurdie and Holmes 2013) and beta diversity was visualized using non-metric multidimensional scaling (NMDS) ordination. Statistical differences for groups identified in the NMDS ordination plots was determined using PERMANOVA (Anderson 2001) using a wrapper function (pairwise.adonis, Martinez

2020) for the adonis function (vegan package version 2.5.6, Oksanen et al. 2016).

Adjusted p-values were reported (Bonferroni correction). Statistical differences between Shannon diversity boxplots was determined using the Wilcoxon rank-sum test (Mann-Whitney, p-values Bonferroni corrected).

In order to identify unique taxa belonging to different strains, the presence/absence of the top 150 ASVs in all samples (which made up 75% of the total counts) were determined for different sample types (seawater, *Oocystis sp.*, *Desmodesmus sp.*) and visualized using a web tool created by VIB/UGent Bioinformatics & Evolutionary Genomics (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Summary taxa plots and alpha diversity boxplots were created using the phyloseq package version 1.28.0 in R (McMurdie and Holmes 2013) as was the abundance heatmap (plot_heatmap function) which implemented the non-clustering NeatMap approach (Rajaram and Oono 2010). The linear discriminant analysis effect size (LEfSe) tool (Segata et al. 2011) was run on the Galaxy platform and used to identify bacterial biomarkers for each strain. The Log₂ fold-change of the 150 most abundant ASVs was calculated from pond inoculation (day 0) to the last day of each pond run (harvested day) and visualized using a heatmap in R (heatmap.2 function, gplots package version 3.0.3).

Stepwise selection of environmental variables and variance partitioning analysis based on Bray Curtis dissimilarity was done in R (vegan package version 2.5.6, Oksanen et al. 2016). Variables tested for significance were strain of algae grown, yearday, age of

culture (days), optical density (440 nm), quantum efficiency of photosystem II (Fv/Fm), percent dissolved oxygen saturation, mean pond temperature (mean from 24 h prior to microbiome sample collection except Day 0 samples), pond temperature range (max temperature – minimum temperature, 24 h prior to microbiome sample collection, range of 0 for Day 0 samples), mean and range of pond pH (both calculated in the same manner as pond temperature mean and range), mean photosynthetically active radiation (PAR, daily mean), mean wind speed (daily), and total daily rainfall.

3. Results

3.1 Microalgae Growth in the Cultivation System Environment

The raceway ponds were highly productive and high nutrient environments where the biofuel-relevant algal strain grown was the dominant primary producer. Each strain of algae was produced at different times of the year and cultivation parameters were optimized, thus each strain had different environmental conditions. *Oocystis sp.* (S002) was grown from early June through early July 2018 with mean pond temperatures of $29.59 \pm 2.32^{\circ}\text{C}$ that varied across time with each pond run (Figure 3B). Pond pH was maintained ~ 6.5 to limit predator abundance, but CO_2 was turned off the morning of the pond harvest in order to allow the pH to rise (>7.5) during the final day of growth to encourage natural settling of the algal cells. Operational issues with the CO_2 system caused two pond runs to have $\text{pH} > 9$ at the time of harvest. *Desmodesmus sp.* (C046) was grown from mid-August to early October 2018 and mean pond temperatures of $28.13 \pm 1.88^{\circ}\text{C}$ with variation across time for each pond run (Figure 3A). For this strain, pH was maintained $\sim 7-7.5$, but typically prior to harvest pH was allowed to rise naturally (to $\sim 7.8-8$) by turning the paddle wheel off and allowing algae to naturally drawdown CO_2 , thus increasing pH. As expected for both strains, algal biomass and dissolved oxygen increased over their growth cycle and ponds were typically harvested just as the algae reached stationary phase (Figure 3C-F). For both strains, specific growth rates of algae varied across different pond runs (Figure 4).

Desmodesmus sp. had specific growth rates ranging from 0.131 to 0.559 d⁻¹ with a mean of 0.320 ± 0.103 (Figure 4A) d⁻¹, while the mean specific growth rate for *Oocystis sp.* was generally higher at 0.420 ± 0.129 d⁻¹ and ranged 0.173 to 0.700 d⁻¹ (Figure 4B).

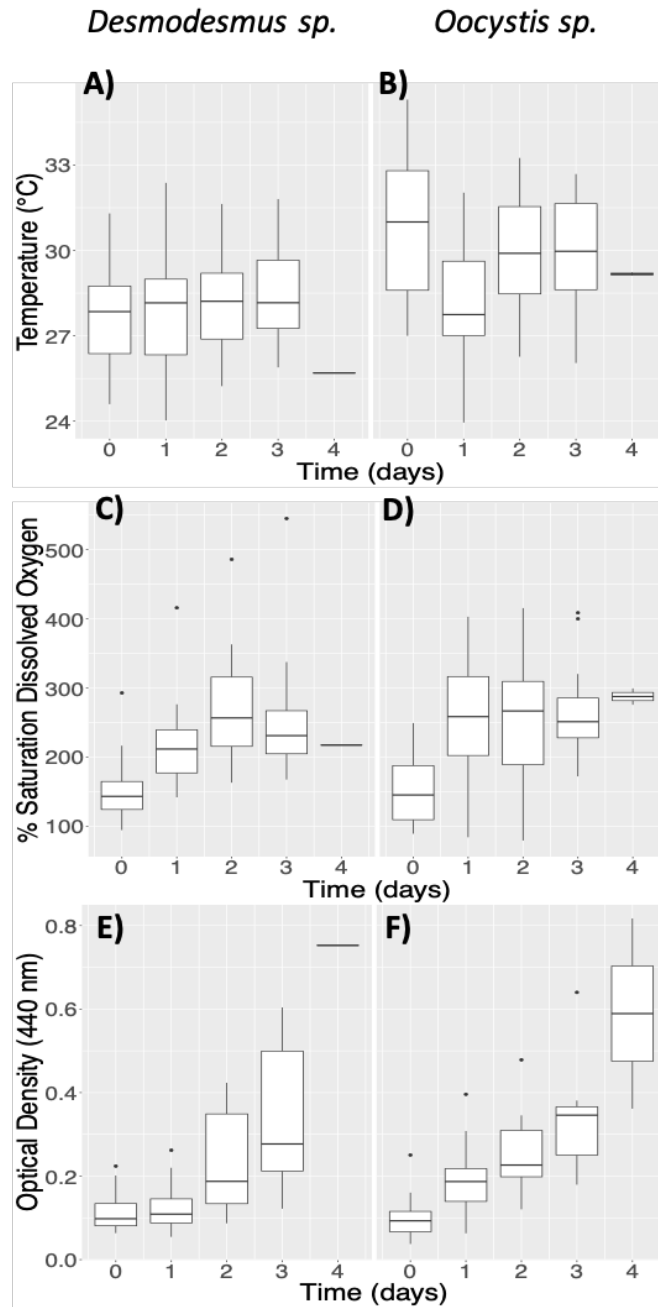


Figure 3: Cultivation environment. Variation in the production pond environment over the course of all pond runs for both *Desmodesmus sp.* (A, C, E) and *Oocystis sp.* (B, D, F). The middle of the boxplot corresponds to the median value and the lower and upper hinges correspond to the first and third quartiles, respectively. Whiskers extend from the hinges to 1.5*IQR (interquartile range) and outliers beyond the whiskers are plotted

individually. A and B) Mean daily temperature (°C), C and D) Optical density (440 nm) of algal biomass, and E and F) Percent saturation dissolved oxygen.

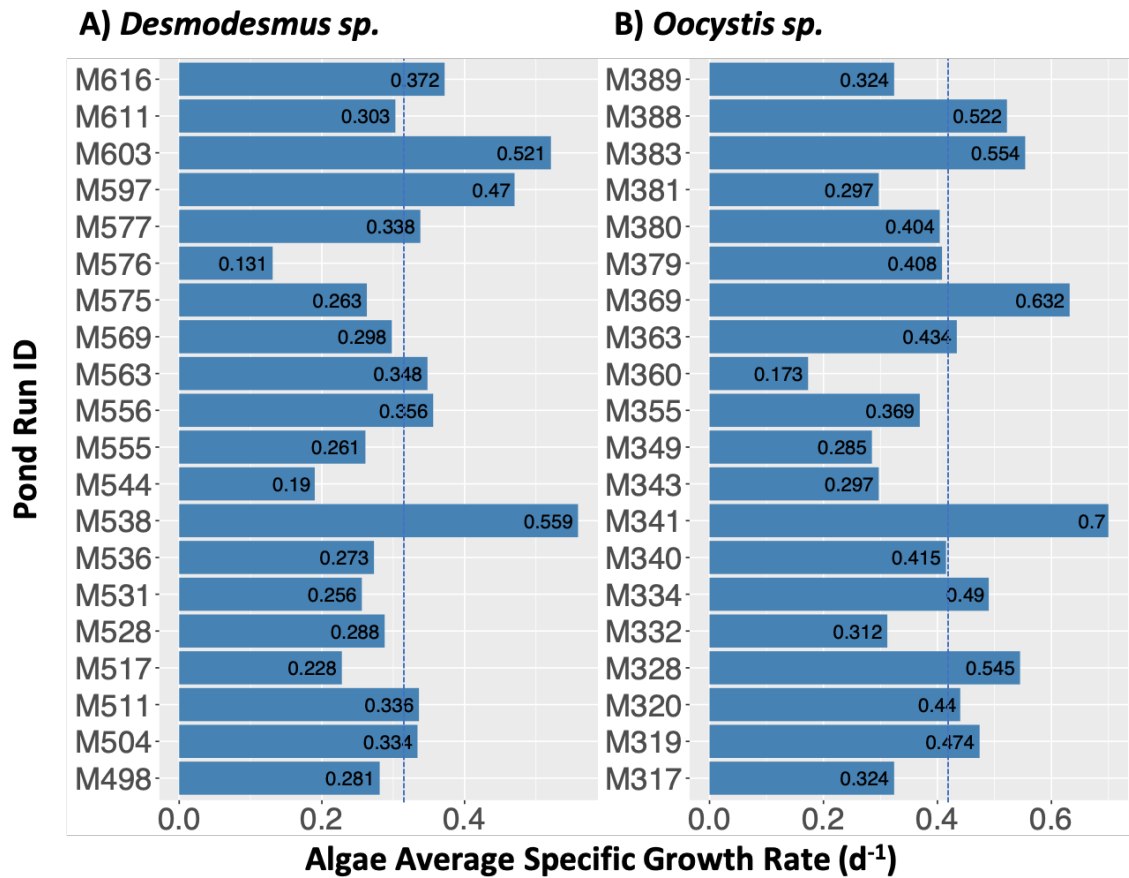


Figure 4: Algae specific growth rate for each pond run. Specific growth rate calculated based on optical density (440 nm) for each pond run when A) *Desmodesmus sp.* (mean $0.320 \pm 0.103 \text{ d}^{-1}$) and B) *Oocystis sp.* (mean $0.420 \pm 0.129 \text{ d}^{-1}$) strains were grown. The dashed lines represent mean specific growth rate for each strain.

3.2 Microbiome Variation Across Host Strains of Algae

At the community level, slight differences between microbiome communities were found for PBRs compared to ponds but there was distinct clustering of microbiomes in the presence of a specific microalgae host (Figure 5). PERMANOVA results found that there was a statistically significant difference between PBR and pond

samples based on Bray Curtis dissimilarity (p-value = 0.024) and both PBR and pond samples were statistically significantly different from treated seawater samples (p-value = 0.006, Figure 5A). Microbiomes from *Oocystis sp.* (S002) samples were statistically different from *Desmodesmus sp.* (C046) samples (p-value = 0.003), and treated seawater samples were statistically different from either strain (p-value = 0.003, Figure 5B). The two outlier communities shown in Figure 5 were identified as unhealthy pond and PBR samples taken when *Oocystis sp.* was grown. The PBR sample was considered to be a crashed sample due to low algae optical density and had been used to inoculate the outlier pond sample (M380D3) three days prior.

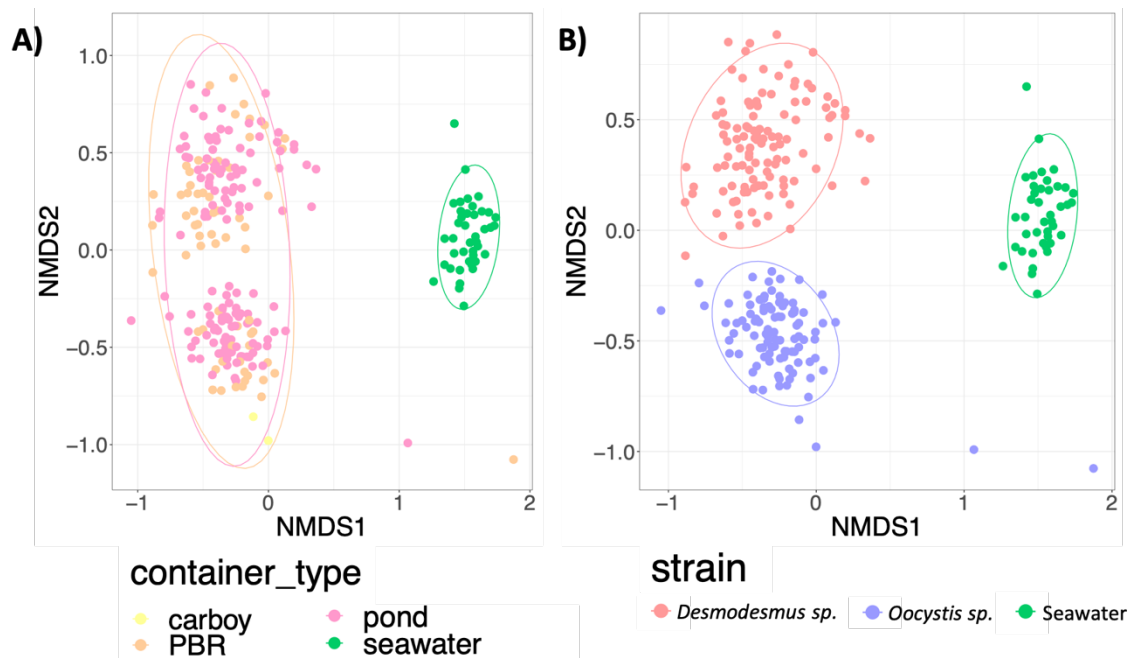


Figure 5: Ordination based on Bray Curtis Dissimilarity. A) NMDS ordination for all microbiome samples grouped by container type. Statistical ellipses are drawn at the 95% confidence level except for carboy samples due to the small sample size (n=2). PERMANOVA results showed significant difference between seawater samples from both pond and PBR samples (p-value = 0.006) and pond and PBR samples were

statistically different from one another (p-value = 0.024). B) NMDS ordination for all microbiome samples grouped by presence of algal strain, *Desmodesmus sp.*, *Oocystis sp.*, or treated seawater inoculum. Statistical ellipses are drawn at the 95% confidence level. PERMANOVA results found significant differences between both *Oocystis sp.* and *Desmodesmus sp.* samples (p-value = 0.003) and significant differences between seawater samples and samples containing either algal strain (p-value = 0.003).

The Shannon diversity of microbial communities in the treated seawater were significantly higher than those found in the samples containing either host strain (Figure 6A). Pairwise comparisons between median Shannon diversity values for samples grown with *Desmodesmus sp.*, *Oocystis sp.*, or treated seawater were all statistically significant from each other ($p < 0.01$) indicated by the pairwise Wilcoxon rank-sum test (Mann-Whitney). Shannon diversity for treated seawater samples was significantly different from either pond or PBR samples ($p < 0.01$, Figure 6B), however pond and PBR diversity was not significantly different ($p = 0.232$, Figure 6B). Carboy samples were excluded from statistical tests due to low number of samples ($n = 2$).

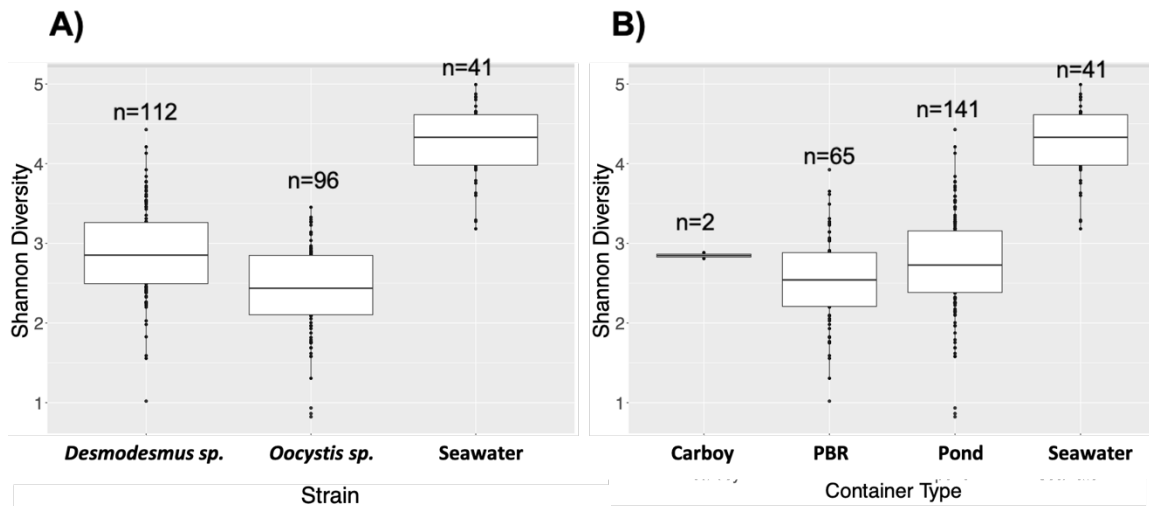


Figure 6: Differences in Shannon diversity across strain and container type. Shannon diversity when compared by A) presence of algal strain (*Desmodesmus sp.*, *Oocystis sp.*, or seawater inoculum) and B) container type the sample was collected from (carboy, PBR, pond, or seawater). The middle of the boxplot corresponds to the median value and the lower and upper hinges correspond to the first and third quartiles, respectively. Whiskers extend from the hinges to 1.5*IQR (interquartile range) and outliers beyond the whiskers are plotted individually. The number of samples used to create each boxplot is indicated by n = X. Pairwise comparisons between both strains and seawater samples were all significantly different ($p < 0.01$). Seawater sample diversity was significantly different from both PBR and pond samples ($p < 0.01$) but pond and PBR diversity were not significantly different ($p = 0.232$).

At the phylotype level, distinctions in the type of bacteria associated with either microalgae strain or the treated seawater inoculum were observed. When considering the top 150 most abundant ASVs (which made up 74% of the total reads), greater than one third were distinct to the treated seawater samples (Figure 7). There were 27 ASVs shared among all sample types and 23 ASVs present when either strain of algae was grown. Only 9 ASVs were unique to each algal strain grown (Figure 7). All microbiome samples analyzed had high relative abundances of bacteria in the *Rhodobacteraceae* family

(~21% of all reads for *Desmodesmus sp.* samples and ~17% of all reads for *Oocystis sp.* samples) however there were differences in the most abundant ASVs depending on the presence of microalgae (Figure 8). Bacteria belonging to the *Saprospiraceae* family had high relative abundances when both algal strains were grown (~42% for *Desmodesmus sp.* and ~21% for *Oocystis sp.*), but bacteria belonging to the *Flavobacteriaceae* and *Cytophagaceae* families had much higher relative abundances when *Oocystis sp.* was grown (~19% and ~21% respectively). ASVs in the *Hyellaceae* and *Hyphomicrobiaceae* families had higher relative abundances when *Desmodesmus sp.* was grown while ASVs in the *Chitinophagaceae*, *Hyphomonadaceae*, and *Verrucomicrobiaceae* families had higher relative abundances when *Oocystis sp.* was grown. Treated seawater samples had much higher relative abundances of *Synechococcaceae*, as well as several other bacterial families that were not present in culture samples (Figure 8). Taxa labeled as “NA” refer to 12 ASVs that could not be classified down to the family level and consisted of 1 *Pseudomonadales* taxon, 1 *Pelagibacterales* taxon, 2 *Oceanospirillales* taxa, 2 *Actinobacteria*, 5 unidentified *Gammaproteobacteria*, and 1 unidentified *Deltaproteobacteria*.

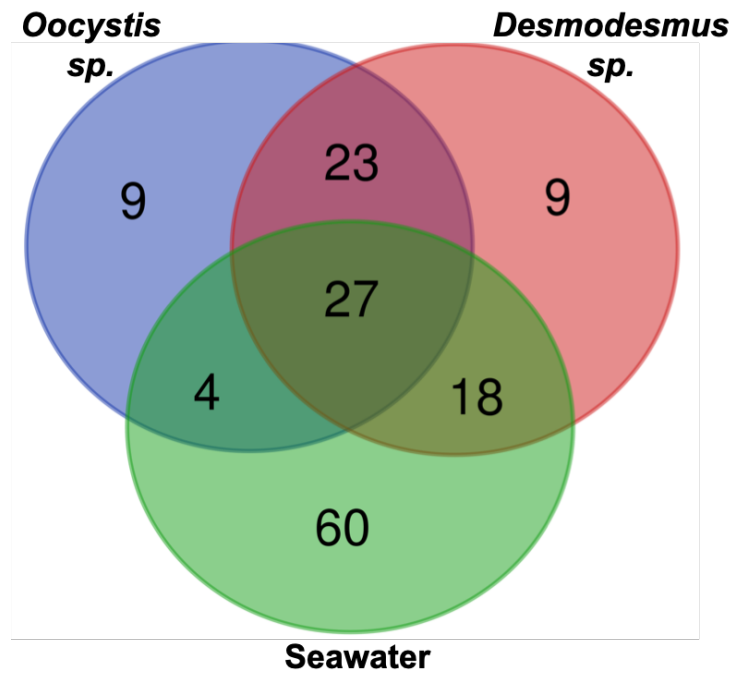


Figure 7: Overlap of the top 150 ASVs between strains. Venn diagram displaying the number of ASVs present (> 5 counts) in culture samples grown with *Oocystis sp.*, *Desmodesmus sp.*, and treated seawater samples for the top 150 most abundant ASVs (74% of total reads).

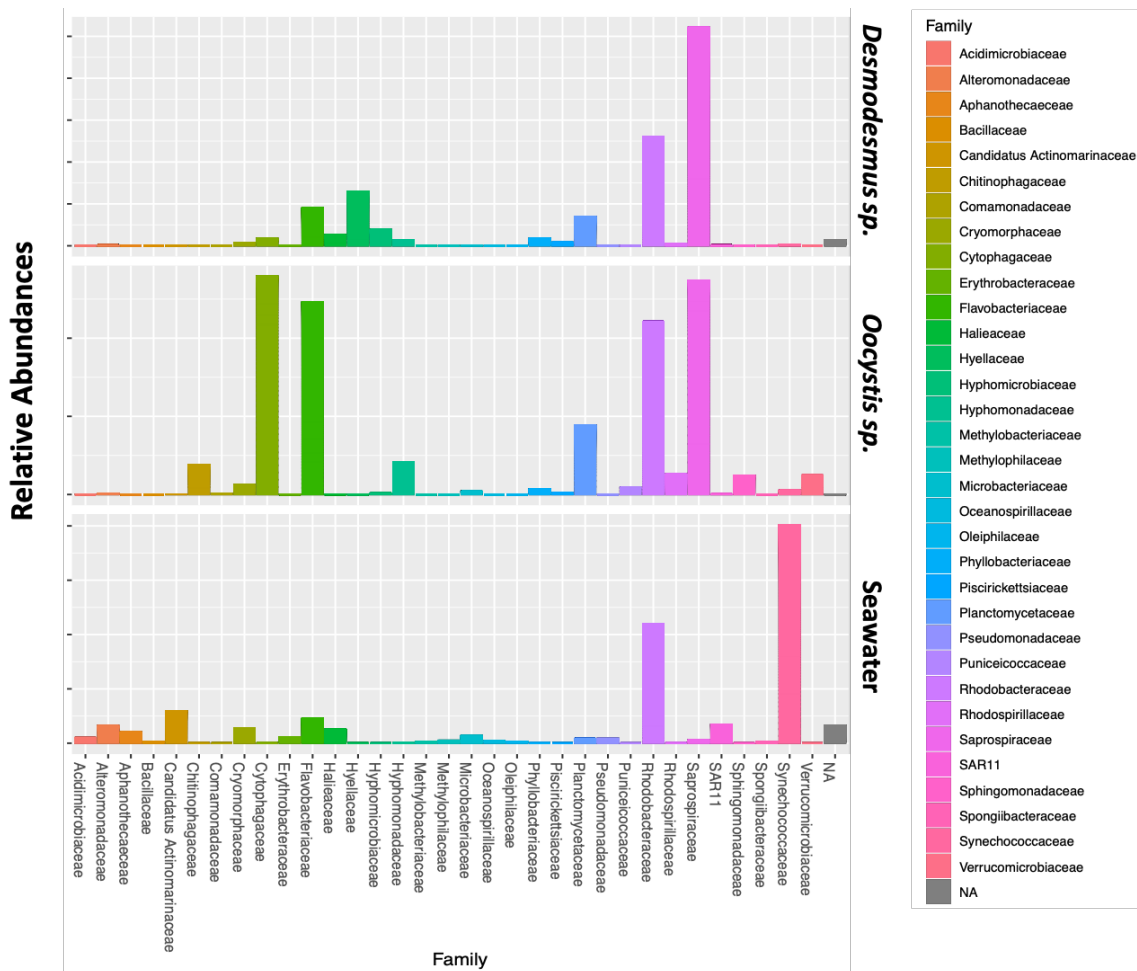


Figure 8: Microbiome variation between algal strains summed at the family level.

Abundance of the top 150 ASVs (74% of total reads) summed across the family level for all samples taken during *Desmodesmus sp.* growth (n = 112), *Oocystis sp.* growth (n = 96) and treated seawater used to inoculate production ponds (seawater, n = 41). Microbiome samples taken when both *Desmodesmus sp.* and *Oocystis sp.* were grown were dominated by *Saprospiraceae* and *Rhodobacteraceae* whereas *Oocystis sp.* also had higher abundances of *Cytophagaceae* and *Flavobacteriaceae*. Samples taken from the treated seawater were dominated by *Synechococcaceae* and had several taxa not detected when algal cells were present. Taxa labeled as NA correspond to 12 ASVs which could not be classified at the family level.

A large number of taxa identified in the treated seawater samples were not detected in production pond, PBR or carboy samples when either algal strain was grown

(Figure 9). These ASVs belong to many bacterial families found in coastal seawater including (but not limited to) *Alteromonadaceae*, *Candidatus Actinomarinaceae*, *Cryomorphaceae*, *Flavobacteriaceae*, *Microbacteriaceae*, SAR11, and *Synechococcaceae*. A cluster of ASVs that had higher relative abundances only in samples containing *Desmodesmus sp.* included *Rhodobacteraceae*, *Planctomycetaceae*, *Flavobacteriaceae*, and *Saprospiraceae* families (Figure 8, Figure 9). Similarly, there was a cluster of ASVs which had much higher relative abundances only in samples containing *Oocystis sp.*, which included members found in the same bacterial families as the *Desmodesmus sp.* abundant ASVs in addition to members of the *Chitinophagaceae*, *Cytophagaceae*, *Hyphomonadaceae*, and *Verrucomicrobiaceae* families (Figure 8, Figure 9). The microbiomes of both algal strains were dominated by some of the same bacterial families (Figure 8), however closely related taxa (ASVs belonging to the same family) showed different preferences to one algal strain over another (Figure 9). Clustering of samples based on container type was not observed as PBR samples clustered closely to the pond samples they were used to seed rather than other PBR samples (Figure 9). The LEfSe analysis identified 13 biomarker ASVs for seawater samples, 12 biomarker ASVs for *Oocystis sp.* samples, and 13 biomarker ASVs for *Desmodesmus sp.* samples (Figure 9). The seawater biomarkers included 3 *Rhodobacteraceae* ASVs and a single ASV belonging to the *Synechococcaceae*, *Alteromonadaceae*, *Pseudomonadaceae*, SAR11, *Microbacteriaceae*, *Haliaceae*, *Erythrobacteraceae*, *Planctomycetaceae*, *Cryomorphaceae*, and *Oceanospirillaceae* families. The

biomarker taxa identified for *Oocystis sp.* samples included 4 *Rhodobacteraceae* ASVs, 2 *Flavobacteriaceae* ASVs, and a single ASV belonging to the *Cytophagaceae*, *Planctomycetaceae*, *Hyphomonadaceae*, *Chitinophagaceae*, *Rhodospirillaceae*, and *Cryomorphaceae* families. The biomarker taxa identified for *Desmodesmus sp.* samples included 4 *Rhodobacteraceae* ASVs, 3 *Saprospiraceae* ASVs, and a single ASV belonging to the *Hyphomicrobiaceae*, *Flavobacteriaceae*, *Planctomycetaceae*, *Oceanospirillaceae*, *Hyphomonadaceae*, and *Piscirickettsiaceae* families.

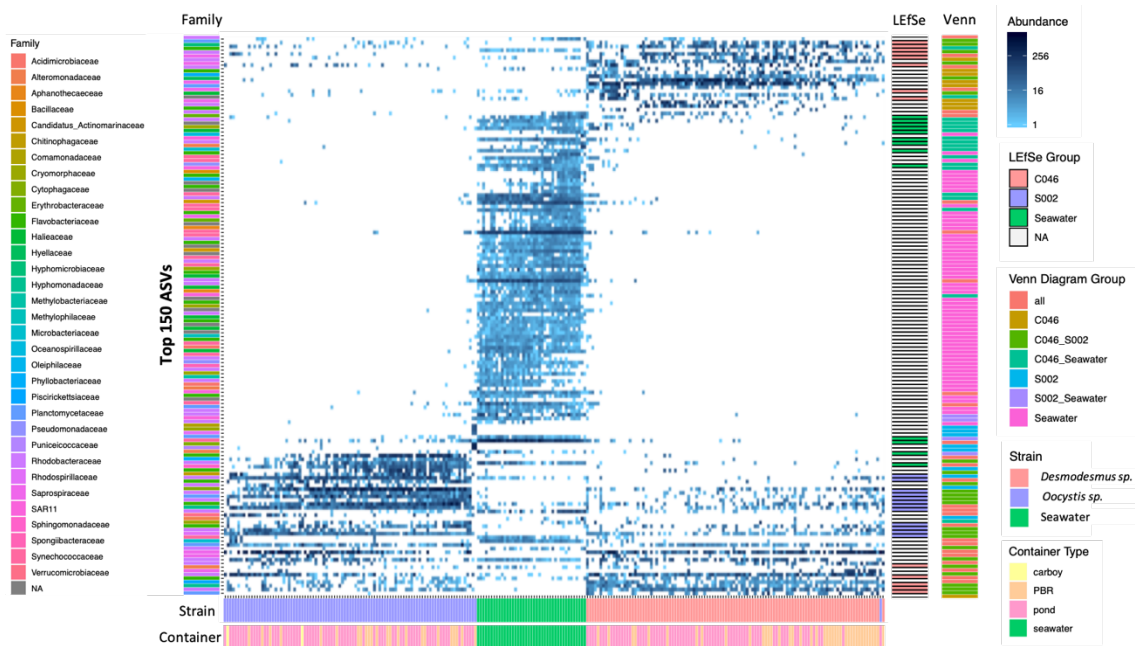


Figure 9: Abundance heatmap for top 150 taxa across all samples. Samples are organized in the heatmap based on NMDS ordination of Bray Curtis dissimilarities with rows corresponding to ASVs and columns corresponding to microbiome samples (NeatMap approach). The Log_4 of rarefied abundance counts is shown with heatmap color, where dark blue represents highest abundance values and white represents lowest values. The color bar along the left y-axis identifies each of the top 150 ASVs at the family level, and taxa labeled as NA could not be classified at the family level. The color bars along the right y-axis show biomarkers identified in the LefSe analysis or category in the Venn diagram (Figure 7). There are two color bars along the x-axis which identify

each sample by strain of algae present (top bar) and container type (bottom bar) the samples were taken from. A large cluster of taxa found primarily in treated seawater samples can be seen along with clusters of taxa that have higher relative abundances in the presence of one algae strain. A cluster of taxa that are present when either algae strain was grown can also be seen and include biomarker taxa identified using LEfSe. PBR samples clustered closer to the pond samples they were used to seed rather than other PBR samples.

3.3 Microbiome Variation Over Time

Microbial communities not only varied between the different container types and different algal host strains, but also over time. Changes in the relative abundance of various taxa were generally consistent between both production ponds (Figure 10). Over the course of each pond run for *Desmodesmus sp.* samples (Day 0 – Day 3) relative abundances of bacterial taxa changed, with distinct shifts in the microbiome community at the start of a new pond run (*Desmodesmus sp.*, Figure 10). A large shift in the dominant taxa was seen in both production ponds for *Oocystis sp.* samples, where bacteria belonging to the families *Cytophagaceae* and *Flavobacteriaceae* had higher relative abundances during the first ~5 pond runs (for both pond 1 and pond 2) and bacteria belonging to the families *Rhodobacteraceae* and *Saprospiraceae* had higher relative abundances for the last ~5 pond runs in each pond (*Oocystis sp.*, Figure 10). ASVs belonging to the family *Verrucomicrobiaceae* were also only seen in later pond runs for *Oocystis sp.* At the community level, diversity typically declined over time during a pond run for *Desmodesmus sp.* (Figure 11B) though there was much variation between pond runs (Figure 11A). The microbiome diversity of *Oocystis sp.* was much more

variable over the course of a pond run (Figure 11) with some pond runs increasing diversity while others decreasing diversity with time.

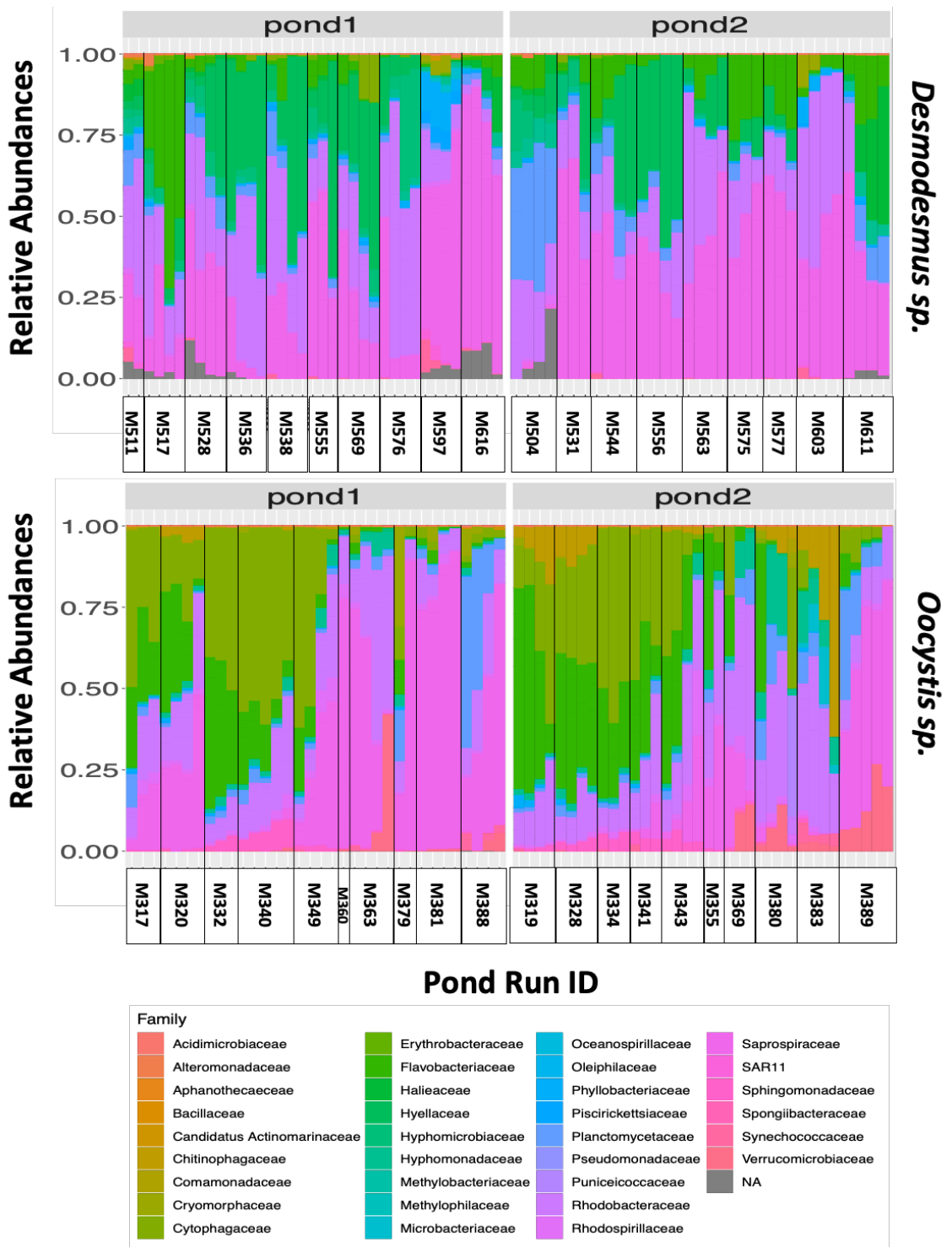


Figure 10: Microbiome variation over time in each production pond during the growth of both algal strains. Relative abundances of the top 150 ASVs (74% of total reads)

classified at the family level over time in the production ponds during the growth of *Desmodesmus sp.* (top panels) and *Oocystis sp.* (bottom panels). Community composition was variable over time, but parallel production ponds (pond 1 and pond 2) were similar.

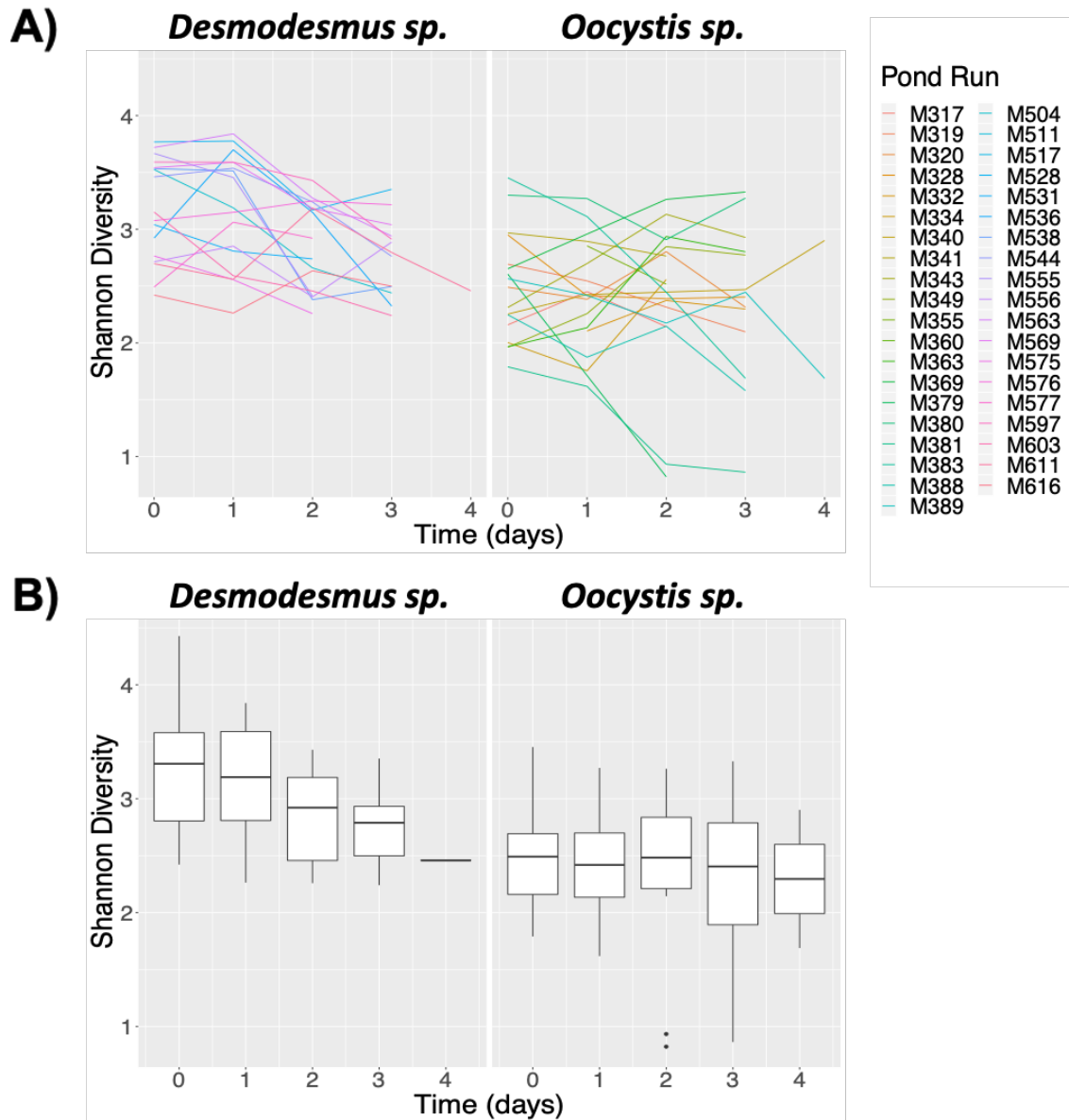


Figure 11: Changes in Shannon diversity over time in production ponds. A) Shannon diversity of each microbiome sample over the course of a pond run for *Desmodesmus sp.* (left panel) and *Oocystis sp.* (right panel). B) Median Shannon diversity for each time point for all pond runs during *Desmodesmus sp.* (left panel) and *Oocystis sp.* (right panel)

growth. A general decline in diversity was observed for *Desmodesmus sp.* whereas diversity was more variable across different pond runs for *Oocystis sp.*

At the ASV level, there was variability in abundance over time between both different algal host strains and different pond runs. Log₂ fold-changes for the top 150 ASVs show that even when the same bacterial taxa were present when either algal strain was grown, their trajectories differed over time (Figure 12). When examining abundance over time for an abundant *Phaeodactylibacter sp.* (ASV_8), it is seen that in the presence of the algal host *Oocystis sp.* it tended to increase in abundance with respect to time, but the opposite was true in the presence of the algal host *Desmodesmus sp.* (Figure 13). Though some taxa generally increased or decreased with time (Figure 13), changes in individual bacterial taxon abundances were variable across all pond runs (Figure 12, Figure 13). Different ASV trajectories were also seen across different pond runs, for example a cluster of five ASVs that had a 4-fold increase during one *Oocystis sp.* pond run (M380) yet that pond run had a similar specific growth rate (0.404 d⁻¹) to the *Oocystis sp.* pond run mean (0.42 d⁻¹, Figure 4). Similarly, the final two *Desmodesmus sp.* pond runs (M611 and M616) had a 5-fold decrease in two ASVs that was less common in other pond runs yet also experienced similar algal specific growth rates (0.303 d⁻¹ and 0.372 d⁻¹, respectively) to the *Desmodesmus sp.* pond run mean (0.32 d⁻¹, Figure 4).

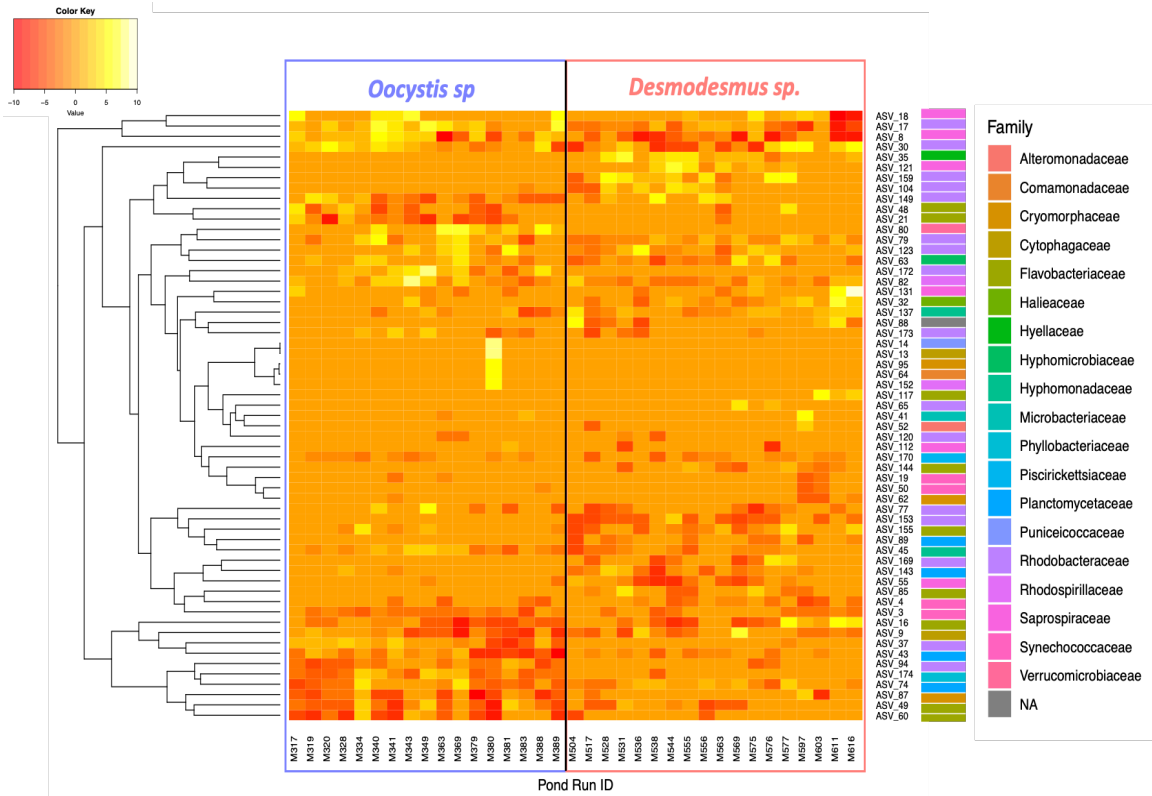


Figure 12: Abundance change for top 150 ASVs during each pond run. Log₂ fold-change for the top 150 ASVs (with abundance > 0 and Log₂ fold-change greater than ±4 in at least one pond run, n = 59) from pond inoculation (Day 0) to the final day (harvest). Fold-change calculations for each pond run are along the x-axis and ordered by date; pond runs when *Oocystis sp.* was grown (indigo outline) and pond runs when *Desmodesmus sp.* was grown (red outline). ASVs are along the y-axis and arranged by hierarchical clustering based on response. Right y-axis represents the family level classification for each ASV.

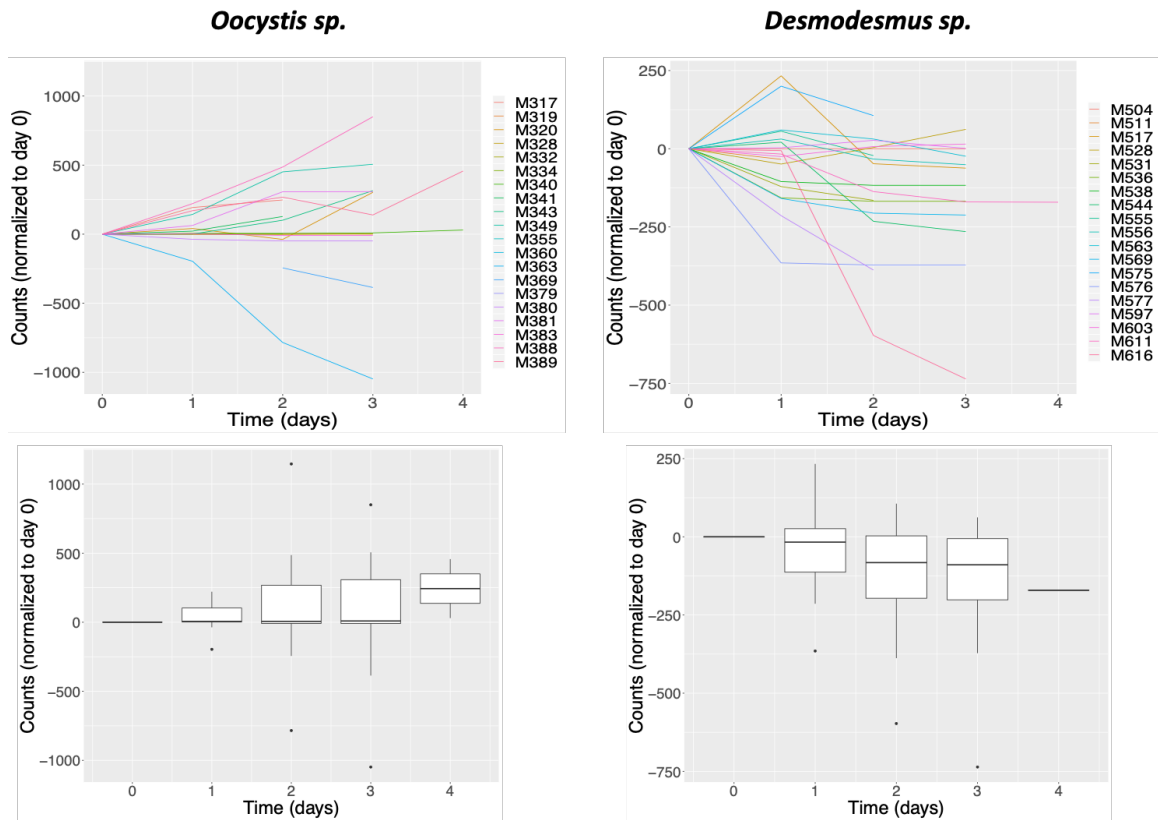


Figure 13: Trajectory for *Phaeodactylibacter sp.* during each pond run. Line plots (top panels) show abundance of a representative *Phaeodactylibacter sp.* (ASV_8, abundance normalized to Day 0) over time for each pond run. Boxplots (bottom panels) are shown for each time point, where the middle of the boxplot corresponds to the median value and the lower and upper hinges correspond to the first and third quartiles, respectively. This *Phaeodactylibacter sp.* generally increased over time when *Oocystis sp.* was grown (left two panels) and generally decreased over time when *Desmodesmus sp.* was grown (right two panels) but trajectory varied for each pond run.

3.4 Environmental Drivers of Microbiome Change

In order to identify important deterministic drivers of microbiome change in the raceway ponds, a stepwise selection of environmental variables was conducted on all production pond samples for both algal strains. Of the metadata variables tested (see Methods), only four were significant: algae strain, yearday, mean pond temperature,

and percent saturation O₂ ($p < 0.01$, Figure 14). Algae strain explained the most variance in pond microbiome samples (19.1%) closely followed by yearday (18.1%). Even though mean pond temperature and percent O₂ saturation were also significant variables, they explained only 3.3% and 1.9% of the variance in the data, respectively. The total variance which could be explained by the four significant environmental variables and their interaction terms was 31.4%, resulting in 68.6% of residual variance unexplained (Table 1).

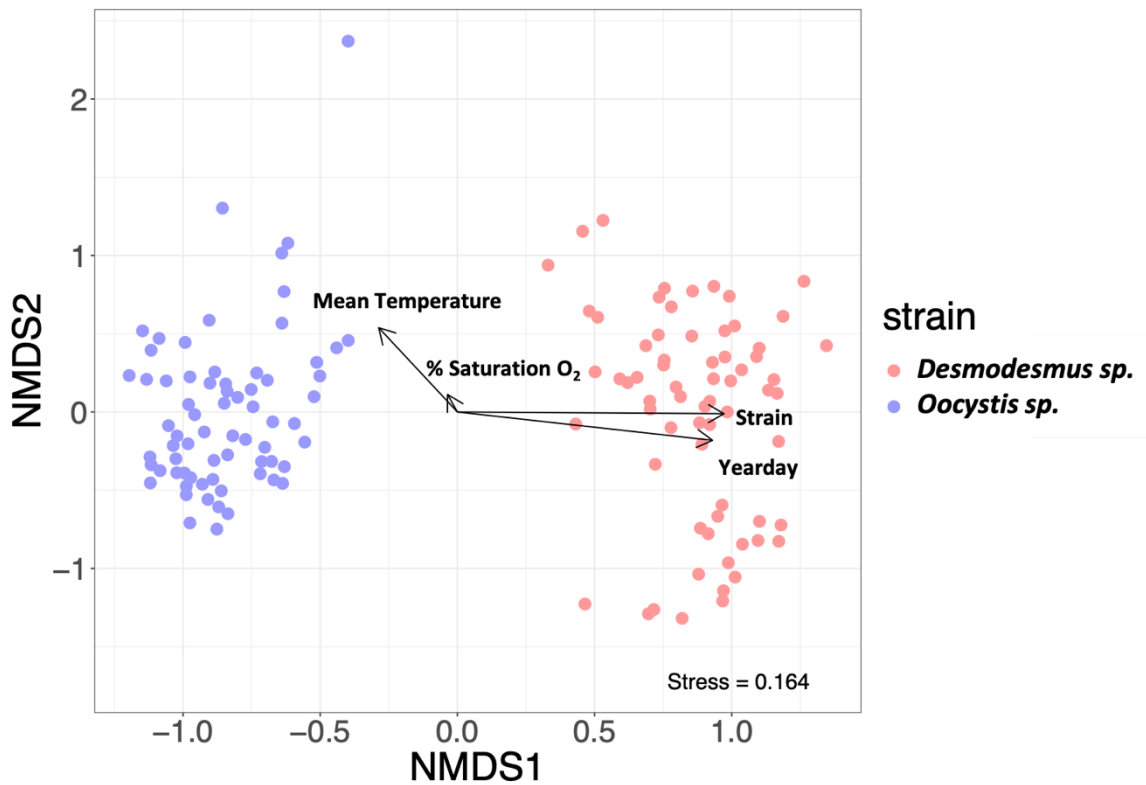


Figure 14: Environmental drivers of microbiome change in production ponds. NMDS ordination plot based on Bray Curtis dissimilarity for production pond samples with arrows corresponding to the four significant metadata variables ($p < 0.01$) identified in the variance partitioning analysis.

Table 1: Variance partitioning analysis on pond environmental drivers. The variance explained by each of the four significant ($p < 0.01$) metadata variables, total variation explained (all four significant variables and their interaction terms, All), and total residual variation.

Variable	% Variance Explained
Strain	19.1%
Yearday	18.5%
Mean Temperature	3.3%
% Saturation O ₂	1.9%
All	31.4%
Residual	68.6%

4. Discussion

The purpose of this study was to identify the microbial community dynamics and their environmental drivers associated with large outdoor marine microalgae raceway ponds. When comparing raceway pond microbiomes to the input communities used to inoculate them, high similarity in diversity and community structure were shown for PBR seed cultures and ponds. Many bacterial taxa that entered the system via the treated seawater inoculum were below detection in pond microbiome samples, likely being outcompeted by algae associated bacteria. This resulted in significantly different communities between pond microbiomes and the seawater inoculum. Similar bacterial families were most abundant when either algal strain was grown, however, specific ASVs tended to associate with either *Desmodesmus sp.* or *Oocystis sp.* Changes in the relative abundances of bacterial taxa with time were observed for all pond runs, however there was no consistent convergence toward the same community structure by the end of each pond run. This study found that the strain of algae cultivated was the most significant driver of microbiome variation which indicates high host selection on the most abundant taxa present. Time, temperature and algal productivity were also important environmental drivers of microbiome community structure.

The cultivation strategy used for this study likely relates to microbial community dynamics, as high inoculation volume and short pond run duration could have led to the similarities between PBR inoculum communities and pond microbiomes observed.

Cultivation for both *Desmodesmus sp.* and *Oocystis sp.* involved seeding the raceway ponds with a high PBR inoculum volume (~25-30%) and short growth period (2-4 days) which has been a successful strategy for large-scale microalgae farms (Huntley et al. 2015). Bacterial communities in the PBRs were preconditioned to the pond environment (high nutrient, high temperature, low pH) and therefore likely had an advantage over the treated seawater community when combined in the production ponds due to priority effects. Priority effects of an inoculum community have been shown to reduce the establishment success of taxa from an invading community (Svoboda et al. 2018), which may be why abundant bacterial taxa in PBR samples continued to be abundant in production pond samples over invading seawater taxa. Production ponds were also completely drained when harvested and re-seeded from PBR cultures, so the priority effects were reset with each iterative pond run. Cultivation strategy and priority effects could explain why pond samples clustered close to the PBRs that were used to inoculate them (Figure 9). If high inoculation volume and short pond run duration lead to microbiomes similar to their seed culture as suggested here, then this could be an effective cultivation strategy when trying to promote an engineered microbiome at larger scales.

The seawater used to inoculate ponds had communities that were significantly different from pond microbiome samples (Figure 5) and had significantly higher diversity (Figure 6). Bacterial families commonly found in the seawater samples (such

as *Synechococcaceae*, *Rhodobacteraceae*, *SAR11*, *Flavobacteriaceae*, *Cryomorphaceae*, *Actinobacteria*, and *Microbacteriaceae*) were consistent with bacterial communities reported for a nearby coastal ocean time series site (Ward et al. 2017), therefore the treatment system did not drastically alter the bacterial community in the seawater source. The large number of bacterial taxa (60) that were highly abundant in the seawater inoculum but were below detection in pond and PBR microbiomes (Figure 7 and Figure 9) were likely quickly outcompeted by bacteria associated with the algal seed cultures. This loss of taxa (and therefore loss in community diversity) could be due to the significant change in environment. The raceway pond environment (Figure 3) is quite shallow (~30 cm deep) and maintains high mean temperatures (27-32°C) and high productivity (100-300 % DO) compared to the coastal ocean environment. Raceway ponds were also maintained at a much lower pH (~6.5-7.5) compared to the coastal ocean environment pH which is ~8-8.1. The treated coastal seawater inoculum was likely able to support a significantly higher bacterial diversity than the pond microbiomes (Figure 6) due to more diverse carbon substrates available in the natural environment. Together, these data suggest that other large algae farms should not necessarily be concerned with removing all bacteria from seawater prior to use in raceway ponds (as is the general practice for lab scale cultivation), which can save time and money for commercialization.

Significant differences in microbiome community structure for *Desmodesmus sp.* pond runs compared to *Oocystis sp.* pond runs were observed (Figure 5). Though there were only nine taxa distinct to each algal strain and overlap of several taxa (Figure 7), the most abundant taxa differed between strains (Figure 9). This is not surprising given that bacterial assemblages attached to algal cells can be algae strain specific (Baker and Kemp 2014). Similarities in the most abundant bacterial families (*Rhodobacteraceae* and *Saprospiraceae*) common between strains may be due to the similar pond environment (high nutrient, high temperature, and high productivity, Figure 3) and that both algal strains were chlorophytes. Dominance of *Saprospiraceae* in raceway ponds was observed and is consistent with algal microbiomes found in other industrial systems (Fulbright et al. 2018). Algal host selection on microbiomes dominated by the bacterial phyla *Proteobacteria* and *Bacteroidetes* have also been seen in diatom laboratory cultures (Amin et al. 2012) and in cultures of other biofuel-relevant strains (Kimbrel et al. 2019, Mark Ibekwe et al. 2017). *Planctomycetes* were also present in the microbiomes of both algal strains and are commonly found in association with microalgae (Lage and Bondoso 2014) as were *Verrucomicrobia* which have been observed in algae raceway ponds growing another green algae *Microchloropsis salina* (Kimbrel et al. 2019).

Desmodesmus sp. microbiomes were significantly more diverse than *Oocystis sp.* microbiomes (Figure 6), however diversity declined with time during *Desmodesmus sp.* pond runs toward a median diversity that was more consistent with *Oocystis sp.* pond

runs over time (Figure 11). There was a greater overlap in taxa between treated seawater inoculum and *Desmodesmus sp.* microbiome samples (Figure 7), so the host selection may have had a delayed affect for *Desmodesmus sp.* compared to *Oocystis sp.* Diversity was more variable with time for *Oocystis sp.* pond runs (Figure 11), which could be related to the different dominant taxa seen for earlier pond runs compared to later pond runs (Figure 10). Differences in quantity and quality of dissolved organic carbon (DOC) are known to play an important role in determining marine bacterial communities (Sarmiento et al. 2016) and could be an important driver for host selection observed in this study, though DOC was not evaluated here.

A shift in the dominant microbiome taxa for the first half of the *Oocystis sp.* pond runs compared to the later pond runs was observed in both pond 1 and pond 2 (Figure 10) and is hypothesized to be driven by lysed algae cells. Visual evidence of the presence of rotifers was confirmed via microscope for several of the PBRs used to inoculate early *Oocystis sp.* pond runs. Marine rotifers are known predators of microalgae and can cause raceway pond cultures to crash (White and Ryan 2015). The first few *Oocystis sp.* pond runs were dominated by *Cytophagaceae* and *Flavobacteriaceae* bacteria when contaminating rotifers were noted while later *Oocystis sp.* pond runs had fewer contamination issues and were dominated by *Rhodobacteraceae* and *Saprospiraceae*. Bacteria belonging to the *Cytophagaceae* family have been known to cause cell lysis in green algae (Cole 1982) and could be another cause of concern when they are present in

high relative abundances in raceway ponds. Early *Oocystis sp.* ponds were more yellow in color and could have been driven by high cell lysis of the algal host due to predators or harmful bacteria. *Rhodobacteraceae* bacteria have been shown to aid in protecting green algae from marine rotifer grazing (Fisher et al. 2019) and could have had a similar effect in this raceway pond system, as they had greater relative abundances in later *Oocystis sp.* pond runs. Different mechanisms for DOC release (such as mechanical cell lysis by sloppy feeding, viral lysis, or algal exudate secretion) lead to distinct DOC compositions (Ma et al. 2018). The mechanical lysis caused by either predator contamination or high *Cytophagaceae* abundance likely created a different quality of DOC in the early *Oocystis sp.* pond runs than in later pond runs where algal secreted DOC was likely dominant and could be a potential mechanism behind the microbiome shift in the *Oocystis sp.* ponds.

Closely related taxa (ASVs that belong to the same family) showed different preferences for algal host strain as there was little clustering of ASVs based on broad phylogenetic groupings (Figure 9). Specific ASVs were sometimes present when either strain was grown, however the relative abundance of that taxa depended on which algae was present. For example, a *Phaeodactylibacter sp.* (ASV_8) was present when either algal strain was grown but was likely better adapted to growing in the presence of *Oocystis sp.*, since its relative abundance increased with time for *Oocystis sp.* but decreased with time for *Desmodesmus sp.* (Figure 13). Taxa level responses were not only variable

between algal hosts, but also variable in their response across different pond runs. For example, the last two pond runs for *Desmodesmus sp.* (M611 and M616) saw a clear decline in ASV_18 and ASV_17 and clear increase in ASV_131 and ASV_32 which were not observed in other *Desmodesmus sp.* pond runs (Figure 12). Another interesting example can be seen for the *Oocystis sp.* pond run M380, where a cluster of five ASVs belonging to different bacterial families (ASV_14, ASV_13, ASV_95, ASV_64, and ASV_152) had a large increase in abundance with time (Figure 12). Interestingly, the final microbiome sample from the M380 pond run before harvesting was also the outlier pond sample identified in Figure 5 and clustered closely to the outlier PBR sample which had been used to seed that pond three days prior. This outlier PBR sample (which was identified as a crashed culture based on optical density) and outlier pond sample were dominated by two ASVs belonging to *Cytophagaceae* and *Puniceicoccaceae* families (ASV_13 and ASV_14, respectively). *Cytophagaceae* are known to cause cell lysis in green algae (Cole 1982) and *Puniceicoccaceae* have been isolated in close association with a marine clamworm (Choo et al. 2007) that feed on marine algae among other things. The two ASVs identified here are likely associated with *Oocystis sp.* pond crashes and may offer potential for early detection of pond crashes at industrial scales.

A lack of convergence in microbiome community change (Figure 12) in combination with similar algal specific growth rates (Figure 4) across production pond runs may indicate that alternate stable state communities occurred. Alternate stable

states in community composition is a well-established phenomenon in the marine environment (Sutherland 1974) and replicate microcosms exposed to the same environmental conditions can produce alternate microbial communities (Pagaling et al. 2017). Repeated culturing conditions (high temperature, high nutrient) in combination with stochasticity in inoculum PBR cultures may have led replicate pond microbiomes to alternate stable states. The decline in ASV_18 and ASV_17 and increase in ASV_131 and ASV_32 for the last two *Desmodesmus sp.* pond runs (M611 and M616) did not yield significantly different algae specific growth rates from mean (Figure 4) and could be an example of an alternate stable microbiome from earlier *Desmodesmus sp.* pond runs. Accepting that alternate steady states are likely to occur outdoor raceway ponds is important to consider when engineering microbiomes for industrial use, as multiple engineered microbiomes may yield similar production results.

Significant environmental drivers of microbiome variation for this system were algal host strain, yearday, temperature and percent dissolved oxygen (Figure 14, Table 1). This data is consistent with similar microalgae industrial systems that have also found algal host as a highly deterministic driver of microbiomes (Kimbrel et al. 2019). Yearday was a proxy for time and was likely significant for this data due to its relationship to strain, as the two algal strains were grown at different times (see Methods) and explained variation in the same ordination direction (Figure 14). Yearday may have also been significant due to the later *Desmodesmus sp.* pond runs clustering

together in the same ordination direction as yearday (Figure 14) and supports the hypothesis that there may have been alternate steady states for *Desmodesmus sp.* pond microbiomes at different times of the year. Percent saturation O₂ can be viewed as a proxy for algal productivity and may have been even more significant had there been more pond crashes in the dataset. The outlier *Oocystis sp.* pond sample thought to be a crashed culture (M380D3) was placed in the same ordination direction as percent saturation O₂ (Figure 14). Primary productivity and temperature are important environmental drivers of microbial communities in the natural marine environment (Ward et al. 2017) therefore it is not surprising that they were significant drivers in the raceway ponds also. The high unexplained variance (68%, Table 1) in the microbiome communities includes a combination of other measured and unmeasured deterministic drivers (such as pH, predator abundance, and biological interactions) and unmeasured stochasticity (such as drift and dispersal). Even though the ponds were operated on short timescales (2-5 days), stochastic processes like drift could have occurred in this system. Scum buildup was observed on the pond lining and persisted over the course of all pond runs for a given strain, therefore some bacteria likely persisted in the system for longer timescales (several weeks) which would be necessary for drift to occur. Given the high unexplained variance in these pond microbiomes, stochastic processes could be important ecological drivers and warrant future investigation in raceway pond systems.

Observations in microbiome dynamics for this microalgae raceway pond system have implications for future research in microbiome engineering and commercialization of microalgae for natural products. Understanding the dominant taxa found in each of the biofuel-relevant strains and the dynamics of their microbiomes in raceway ponds can be used as a baseline for future microbiome manipulation experiments. This study shows that cultivation strategy, such as inoculation volume and cultivation period, could be important for microbiome consistency between seed culture and larger ponds which will be critical for promoting the survival of an engineered seed community. The observations in this study also suggest the importance of growing a seed inoculum at similar growth conditions (high temperature, high productivity) to help ensure priority effects over non-sterile seawater community members. This cultivation system focused on treating seawater to remove predators or weed algae (Figure 1) rather than rendering it completely sterile, which was still an effective treatment system for raceway ponds as seawater associated bacteria were significantly less abundant.

Because of the high variability in a particular ASV's trajectory with time and the stochasticity likely in this system, future co-culturing experiments between industrially relevant algae and bacteria should focus on engineering a microbial consortium rather than an axenic co-culture. Additions of individual bacterial strains into microalgae cultures with an existing microbiome have shown lasting effects (Baker and Kemp 2020) and may be an effective method for developing effective microbial consortia. Fisher et

al. (2019) found a microbial consortium that helped protect a host microalga from rotifer predation, but that protective effect was lost in the absence of occasional contamination. Therefore, maintaining several effective microbiome communities (perhaps alternate steady state communities) may be an effective microbiome engineering approach for algae cultivation at large scales.

Valuable lessons can be learned from observational work discussed here, however there were also many limitations in this study. The microbiomes of only two different algal strains, which were both green algae, were tested in this work and information on bacteria associated with different types of biofuel-relevant algae would be useful. Also, more microbiome samples from the carboy (indoor laboratory cultures) would have been beneficial for confirming specific algae-associated bacteria for both strains. This study only considered the bacterial component of the pond microbiome and didn't include the broader microbial community such as viruses and algal predators. The viral community could have an important bottom up effect on the pond microbiome and would be critical to consider when engineering seed communities. Predator contamination was noted but not quantified and would have been important for confirming potential harmful or protective bacterial taxa. All microbiome samples taken in this study were from ponds grown during summer and early fall, and different seasons may cause different environmental drivers to be significant. The quantity and

quality of DOC was not measured for this system and was hypothesized to play an important role in determining the microbiome community.

5. Conclusion

This study described the microbiome dynamics and environmental drivers of community change in large-scale raceway ponds during the growth of two biofuel-relevant marine microalgae strains, *Desmodesmus* sp. (C046) and *Oocystis* sp. (S002). Pond microbiome communities were significantly different from input seawater communities but similar in diversity and community structure to PBR seed cultures. Cultivation strategy, namely high inoculation volumes and short cultivation periods, may have created priority effects for seed culture bacterial communities to prevail over invading seawater bacteria, causing them to play a less important role in pond microbiome dynamics. The algae strain (host) had high selection effects on the most abundant taxa present in ponds, however some bacterial families were present when either strain was grown. Contamination by predators or lytic bacteria can cause drastic changes in the pond raceway microbiome, from healthy cultures dominated by *Rhodobacteraceae* and *Saprospiraceae* to unhealthy cultures dominated by *Cytophagaceae* and *Puniceicoccaceae*. Generally, closely related taxa did not consistently change with time in the raceway ponds and a high variability in ASV fold-change was observed between pond runs. The algal strain (host) grown was the most dominant driver of microbiome variation in the production ponds, followed by changes in time and changes in the natural environment (i.e. temperature and productivity).

Thinking forward, increasing the consistency and efficiency of growing marine microalgae at large industrial scales is essential for algae biofuels to be economically feasible. Consistency in biomass production, through reduced crop loss via unhealthy pond cultures, can potentially be improved when considering the microbiome community in cultivation plans. Focusing future research on engineering multiple robust microbial consortiums rather than trying to maintain optimal algae-bacteria co-cultures will be more effective when scaling up cultivation to larger raceway ponds, as they are diverse and variable environments. Though raceway pond diversity is significantly reduced from communities in the natural environment, outdoor ponds will still have greater diversity than typically found laboratory cultures and could be advantageous for culture resilience.

Successful microbiome engineering for outdoor raceway ponds will require future research on interactions between the entire microbial community in these systems, including viruses, bacteria, algae, fungi, and protists. Early detection of potential threats to algae productivity would be important in a commercial setting and could inform the design of engineered seed communities or cultivation strategies. Rapidly dropping pond pH to reduce predator abundance when growing algae at larger scales (Ganuza et al. 2016) is an example of a cultivation strategy that could have significant impacts on the pond microbiome. Future testing of engineered microbiome

communities in raceway ponds would be most effective at semi-industrial scales in order to test for effectiveness and community divergence while still being cost effective.

Further analysis of this data may examine potential functional responses of different communities by using a predictive metagenome software program such as PICRUSt (Langille et al. 2013). Testing for functional differences between distinct communities such as between algal strain or healthy compared to unhealthy ponds could warrant interesting results. A predicted metagenome would also be insightful for determining whether alternate steady state communities emerged in the ponds. Future laboratory scale experiments could be conducted to test for priority effects and alternate steady state communities hypothesized in this analysis. Treated seawater communities could be inoculated with varying volumes of algae seed culture to determine the minimum percentage of inoculation culture needed to maintain potential priority effects. Several replicate *Desmodesmus sp.* and *Oocystis sp.* cultures could be inoculated from a single source culture and maintained at constant growth conditions to see if different steady state communities emerge and if any are similar in community composition to those observed in the outdoor production ponds.

The field of algae microbiome science will continue to develop as interests in commercialization of marine microalgae for natural products persists. Advances in modeling of microbial communities to determine potential effective consortiums prior to testing them at larger scales could be an efficient method in this field. Utilization of

bacterial consortia for wastewater treatment is a well-developed process and there may be more interest in incorporating algae ponds into these systems in the future. Overall there is much left to learn about utilizing microbial consortia for improving marine microalgae cultivation strategies but will hopefully lead sustainable biofuel sources in the future.

References

- Amin, S. A., Parker, M. S., & Armbrust, E. V. (2012). Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews*, 76(3), 667–684. doi: 10.1128/MMBR.00007-12.
- Andersen, R. A., Berges, J. A., Harrison, P. J., & Watanabe, M. M. (2005). Appendix A - Recipes for Freshwater and Seawater Media. In *Algal Culturing Techniques* (pp. 429–538). Elsevier. doi: 10.1016/B978-012088426-1/50027-5.
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance: NON-PARAMETRIC MANOVA FOR ECOLOGY. *Austral Ecology*, 26(1), 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x.
- Baker, L. J., & Kemp, P. F. (2020). Bacterial inoculations can perturb the growth trajectory of diatoms with an existing microbiome. *PeerJ*, 8, e8352. doi: 10.7717/peerj.8352.
- Baker, L., & Kemp, P. (2014). Exploring bacteria–diatom associations using single-cell whole genome amplification. *Aquatic Microbial Ecology*, 72(1), 73–88. doi: 10.3354/ame01686.
- Behrenfeld, M. J., & Falkowski, P. G. (1997). A consumer’s guide to phytoplankton primary productivity models. *Limnology and Oceanography*, 42(7), 1479–1491. doi: 10.4319/lo.1997.42.7.1479.
- Beier, S., Rivers, A. R., Moran, M. A., & Obernosterer, I. (2015). The transcriptional response of prokaryotes to phytoplankton-derived dissolved organic matter in seawater: The response of prokaryotes to DOM in seawater. *Environmental Microbiology*, 17(10), 3466–3480. doi: 10.1111/1462-2920.12434.
- Bell, W., & Mitchell, R. (1972). Chemotactic and growth responses of marine bacteria to algal extracellular products. *The Biological Bulletin*, 143(2), 265–277. doi: 10.2307/1540052.
- Bhujade, R., Chidambaram, M., Kumar, A., & Sapre, A. (2017). Algae to Economically Viable Low-Carbon-Footprint Oil. *Annual Review of Chemical and Biomolecular Engineering*, 8(1), 335–357. doi: 10.1146/annurev-chembioeng-060816-101630.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. doi: 10.1038/nmeth.3869.
- Chase, J. M., & Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1576), 2351–2363. doi: 10.1098/rstb.2011.0063.

- Chase, Jonathan M. (2003). Community assembly: When should history matter? *Oecologia*, 136(4), 489–498. doi: 10.1007/s00442-003-1311-7.
- Choo, Y.-J., Lee, K., Song, J., & Cho, J.-C. (2007). Puniceicoccus vermicola gen. Nov., sp. Nov., a novel marine bacterium, and description of Puniceicoccaceae fam. Nov., Puniceicoccales ord. Nov., Opitutaceae fam. Nov., Opitutaes ord. Nov. And Opitutae classis nov. In the phylum 'Verrucomicrobia.' *International Journal of Systematic and Evolutionary Microbiology*, 57(3), 532–537. doi: 10.1099/ijs.0.64616-0.
- Coder, D. M., & Starr, M. P. (1978). Antagonistic association of the chlorellavirus bacterium ("Bdellovibrio" chlorellavirus) with *Chlorella vulgaris*. *Current Microbiology*, 1(1), 59–64. doi: 10.1007/BF02601710.
- Cole, J. J. (1982). Interactions Between Bacteria and Algae in Aquatic Ecosystems. *Annual Review of Ecology and Systematics*, 13(1), 291–314. doi: 10.1146/annurev.es.13.110182.001451.
- Fisher, C. L., Ward, C. S., Lane, P. D., Kimbrel, J. A., Sale, K. L., Stuart, R. K., Mayali, X., & Lane, T. W. (2019). Bacterial communities protect the alga *Microchloropsis salina* from grazing by the rotifer *Brachionus plicatilis*. *Algal Research*, 40, 101500. doi: 10.1016/j.algal.2019.101500.
- Foster, R. A., Kuypers, M. M. M., Vagner, T., Paerl, R. W., Musat, N., & Zehr, J. P. (2011). Nitrogen fixation and transfer in open ocean diatom–cyanobacterial symbioses. *The ISME Journal*, 5(9), 1484–1493. doi: 10.1038/ismej.2011.26.
- Fukami, T. (2015). Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 1–23. doi: 10.1146/annurev-ecolsys-110411-160340.
- Fullbright, S. P., Robbins-Pianka, A., Berg-Lyons, D., Knight, R., Reardon, K. F., & Chisholm, S. T. (2018). Bacterial community changes in an industrial algae production system. *Algal Research*, 31, 147–156. doi: 10.1016/j.algal.2017.09.010.
- Ganuja, E., Sellers, C. E., Bennett, B. W., Lyons, E. M., & Carney, L. T. (2016). A Novel Treatment Protects *Chlorella* at Commercial Scale from the Predatory Bacterium *Vampirovibrio chlorellavorus*. *Frontiers in Microbiology*, 7. doi: 10.3389/fmicb.2016.00848.
- Herrera Paredes, S., Gao, T., Law, T. F., Finkel, O. M., Mucyn, T., Teixeira, P. J. P. L., Salas González, I., Feltcher, M. E., Powers, M. J., Shank, E. A., Jones, C. D., Jojic, V., Dangl, J. L., & Castrillo, G. (2018). Design of synthetic bacterial communities for predictable plant phenotypes. *PLOS Biology*, 16(2), e2003962. doi: 10.1371/journal.pbio.2003962.

- Huntley, M. E., Johnson, Z. I., Brown, S. L., Sills, D. L., Gerber, L., Archibald, I., Machesky, S. C., Granados, J., Beal, C., & Greene, C. H. (2015). Demonstrated large-scale production of marine microalgae for fuels and feed. *Algal Research*, *10*, 249–265. doi: 10.1016/j.algal.2015.04.016.
- Johnson, Z. (2004). Description and application of the background irradiance gradient-single turnover fluorometer (BIG-STf). *Marine Ecology Progress Series*, *283*, 73–80. doi: 10.3354/meps283073.
- Johnson, Z. I., Shyam, R., Ritchie, A. E., Mioni, C., Lance, V. P., Murray, J. W., & Zinser, E. R. (2010). The effect of iron- and light-limitation on phytoplankton communities of deep chlorophyll maxima of the western Pacific Ocean. *Journal of Marine Research*, *68*(2), 283–308. doi: 10.1357/002224010793721433.
- Jutson, M. G. S., Pipe, R. K., & Tomas, C. R. (2016). The Cultivation of Marine Phytoplankton. In *Microalgae: Current Research and Applications* (pp. 11–26). Caister Academic Press. doi: 10.21775/9781910190272.02.
- Kazamia, E., Czesnick, H., Nguyen, T. T. V., Croft, M. T., Sherwood, E., Sasso, S., Hodson, S. J., Warren, M. J., & Smith, A. G. (2012). Mutualistic interactions between vitamin B12-dependent algae and heterotrophic bacteria exhibit regulation: Algal-bacterial interactions for delivery of vitamin B12. *Environmental Microbiology*, *14*(6), 1466–1476. doi: 10.1111/j.1462-2920.2012.02733.x.
- Kimbrel, J. A., Samo, T. J., Ward, C., Nilson, D., Thelen, M. P., Siccardi, A., Zimba, P., Lane, T. W., & Mayali, X. (2019). Host selection and stochastic effects influence bacterial community assembly on the microalgal phycosphere. *Algal Research*, *40*, 101489. doi: 10.1016/j.algal.2019.101489.
- Lage, O. M., & Bondoso, J. (2014). Planctomycetes and macroalgae, a striking association. *Frontiers in Microbiology*, *5*. doi: 10.3389/fmicb.2014.00267.
- Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkepile, D. E., Vega Thurber, R. L., Knight, R., Beiko, R. G., & Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814–821. doi: 10.1038/nbt.2676.
- Larkin, A. A., Blinebry, S. K., Howes, C., Lin, Y., Loftus, S. E., Schmaus, C. A., Zinser, E. R., & Johnson, Z. I. (2016). Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. *The ISME Journal*, *10*(7), 1555–1567. doi: 10.1038/ismej.2015.244.

- Lindemann, S. R., Bernstein, H. C., Song, H.-S., Fredrickson, J. K., Fields, M. W., Shou, W., Johnson, D. R., & Beliaev, A. S. (2016). Engineering microbial consortia for controllable outputs. *The ISME Journal*, *10*(9), 2077–2084. doi: 10.1038/ismej.2016.26.
- Lozupone, C. A., & Knight, R. (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences*, *104*(27), 11436–11440. doi: 10.1073/pnas.0611525104.
- Luo, S., Waller, L., Badgley, B., He, Z., & Young, E. B. (2020). Effects of bacterial inoculation and nitrogen loading on bacterial-algal consortium composition and functions in an integrated photobioelectrochemical system. *Science of The Total Environment*, *716*, 137135. doi: 10.1016/j.scitotenv.2020.137135.
- Ma, X., Coleman, M. L., & Waldbauer, J. R. (2018). Distinct molecular signatures in dissolved organic matter produced by viral lysis of marine cyanobacteria: Molecular signatures of viral DOM release. *Environmental Microbiology*, *20*(8), 3001–3011. doi: 10.1111/1462-2920.14338.
- Mark Ibekwe, A., Murinda, S. E., Murry, M. A., Schwartz, G., & Lundquist, T. (2017). Microbial community structures in high rate algae ponds for bioconversion of agricultural wastes from livestock industry for feed production. *Science of The Total Environment*, *580*, 1185–1196. doi: 10.1016/j.scitotenv.2016.12.076.
- Martinez, A. (2020). *pairwiseAdonis: Pairwise multilevel comparison using adonis* (Version 0.4) [R package].
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. doi: 10.1371/journal.pone.0061217.
- Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. *MBio*, *3*(2). doi: 10.1128/mBio.00036-12.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package* (Version 2.5-6) [R package]. <https://CRAN.R-project.org/package=vegan>.
- Pagaling, E., Vassileva, K., Mills, C. G., Bush, T., Blythe, R. A., Schwarz-Linek, J., Strathdee, F., Allen, R. J., & Free, A. (2017). Assembly of microbial communities in replicate nutrient-cycling model ecosystems follows divergent trajectories, leading to alternate stable states: Divergent trajectories in microbial communities. *Environmental Microbiology*, *19*(8), 3374–3386. doi: 10.1111/1462-2920.13849.
- Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time

- series and global field samples: Primers for marine microbiome studies. *Environmental Microbiology*, 18(5), 1403–1414. doi: 10.1111/1462-2920.13023.
- Rajaram, S., & Oono, Y. (2010). NeatMap—Non-clustering heat map alternatives in R. *BMC Bioinformatics*, 11(1), 45. doi: 10.1186/1471-2105-11-45.
- Ritchie, R. J. (2006). Consistent Sets of Spectrophotometric Chlorophyll Equations for Acetone, Methanol and Ethanol Solvents. *Photosynthesis Research*, 89(1), 27–41. doi: 10.1007/s11120-006-9065-9.
- Rivett, D. W., Jones, M. L., Ramoneda, J., Mombrikotb, S. B., Ransome, E., & Bell, T. (2018). Elevated success of multispecies bacterial invasions impacts community composition during ecological succession. *Ecology Letters*, 21(4), 516–524. doi: 10.1111/ele.12916.
- Rummens, K., De Meester, L., & Souffreau, C. (2018). Inoculation history affects community composition in experimental freshwater bacterioplankton communities: Experimental priority effects in bacterioplankton. *Environmental Microbiology*, 20(3), 1120–1133. doi: 10.1111/1462-2920.14053.
- Sarmiento, H., & Gasol, J. M. (2012). Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton: Use of phytoplankton-derived DOC by bacterioplankton. *Environmental Microbiology*, 14(9), 2348–2360. doi: 10.1111/j.1462-2920.2012.02787.x.
- Sarmiento, H., Morana, C., & Gasol, J. M. (2016). Bacterioplankton niche partitioning in the use of phytoplankton-derived dissolved organic carbon: Quantity is more important than quality. *The ISME Journal*, 10(11), 2582–2592. doi: 10.1038/ismej.2016.66.
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, 12(6), R60. <https://doi.org/10.1186/gb-2011-12-6-r60>
- Seyedsayamdost, M. R., Case, R. J., Kolter, R., & Clardy, J. (2011). The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nature Chemistry*, 3(4), 331–335. doi: 10.1038/nchem.1002.
- Sutherland, J. P. (1974). Multiple Stable Points in Natural Communities. *The American Naturalist*, 108(964), 859–873. doi: 10.1086/282961.
- Svoboda, P., Lindström, E. S., Ahmed Osman, O., & Langenheder, S. (2018). Dispersal timing determines the importance of priority effects in bacterial communities. *The ISME Journal*, 12(2), 644–646. doi: 10.1038/ismej.2017.180.

- Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology*, 73(16), 5261–5267. doi: 10.1128/AEM.00062-07.
- Wang, X., Li, Z., Su, J., Tian, Y., Ning, X., Hong, H., & Zheng, T. (2010). Lysis of a red-tide causing alga, *Alexandrium tamarense*, caused by bacteria from its phycosphere. *Biological Control*, 52(2), 123–130. doi: 10.1016/j.biocontrol.2009.10.004.
- Ward, C. S., Yung, C.-M., Davis, K. M., Blinebry, S. K., Williams, T. C., Johnson, Z. I., & Hunt, D. E. (2017). Annual community patterns are driven by seasonal switching between closely related marine bacteria. *The ISME Journal*, 11(6), 1412–1422. doi: 10.1038/ismej.2017.4.
- White, R. L., & Ryan, R. A. (2015). Long-Term Cultivation of Algae in Open-Raceway Ponds: Lessons from the Field. *Industrial Biotechnology*, 11(4), 213–220. doi: 10.1089/ind.2015.0006.
- Zhou, J., Deng, Y., Zhang, P., Xue, K., Liang, Y., Van Nostrand, J. D., Yang, Y., He, Z., Wu, L., Stahl, D. A., Hazen, T. C., Tiedje, J. M., & Arkin, A. P. (2014). Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proceedings of the National Academy of Sciences*, 111(9), E836–E845. doi: 10.1073/pnas.1324044111.
- Zhou, J., & Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiology and Molecular Biology Reviews*, 81(4), e00002-17, e00002-17. doi: 10.1128/MMBR.00002-17.