IMPACT OF DAYTIME NITROGEN LIMITATION ON STRUCTURE AND FUNCTION OF MIXED MICROALGAL COMMUNITIES DERIVED FROM THREE GEOGRAPHICALLY DISTINCT SOURCES

BY

ANNA CHRISTINE FEDDERS

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Environmental Engineering in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Adviser:

Assistant Professor Jeremy Guest

ABSTRACT

Increased nitrogen (N) and phosphorus (P) loadings to surface waters lead to algal blooms which cause eutrophication and hypoxia, threatening aquatic ecosystems, human health, and local economies. The adoption of increasingly stringent regulations for effluent N and P loadings, including from wastewater treatment plants (WWTPs), is a successful strategy for addressing this challenge. However, existing technologies for N and P removal are inconsistent with sustainability goals for nutrient recovery and reduced energy consumption. Eukaryotic microalgae are uniquely equipped to address these challenges, with the capacity to remove inorganic and organic N and P from growth media to the point of non-detection and to accumulate carbohydrate and lipid stores which may be further converted to biofuels. While algal wastewater treatment and biofuels are not new lines of research, much previous work has focused on either pure cultures or minimally engineered open systems. Algal biotechnology, using highly engineered open systems, seeks a middle road between these two approaches, offering greater control than open ponds but requiring fewer inputs than axenic systems. The work presented in this thesis explores nutrient dynamics, carbon storage, and community dynamics of mixed algal cultures within the context of algal biotechnology with an aim toward filling key knowledge gaps. Three parallel photobioreactors, each inoculated with mixed microbial communities sourced from WWTP clarifiers and/or surface water in one of three different geographic locations of the United States (Florida, North Carolina, Illinois) were operated in sequencing batch mode with an 8-day solids residence time (SRT) and 14hr:10hr light:dark cycle with nighttime N-feeding and daytime N-limitation. Metrics of community function included nutrient dynamics and biomass composition which were monitored over the course of 82 days and during a subsequent 10-day batch experiment to determine kinetic parameters. Community structure over 82 days was monitored via sequencing of the V4 and V8-V9 regions of the 18S rRNA gene. Structural and functional metrics showed complementary patterns. Functional metrics transitioned from dynamic to stable performance through time while community structure showed a departure from, followed by a return to, a community resembling the initial community. Although the communities in the three reactors remained distinct from each other through time, the most dominant OTUs were shared between all three reactors. The data presented in this experiment demonstrate how system design can induce communities which differ in structure to follow similar extant performance patterns. However, intrinsic storage parameters for carbohydrates and lipids differed widely between reactors, suggesting that more research is needed to elucidate structure-function relationships in microalgal communities to maximize carbon storage potential.

ACKNOWLEDGEMENTS

Special thanks to my parents, James and Barbara, and sisters, Rachel and Emily Fedders, for their unconditional love and support.

I am grateful to Jeremy Guest, whose expertise, guidance, patience, generosity, respect, and good humor as an advisor made this research possible. It has been a privilege to have had the opportunity to work under his guidance. Thanks to the other members of the Guest group for providing a supportive working environment and especially to Diana Byrne who it has been my privilege to have as a desk neighbor for three years. I will miss swapping ideas and sharing life chats and chocolate with you. Ian Bradley provided expert guidance and training in algal cultivation, DNA sequencing, and data analysis. Ian, please make sure you leave a copy of your brain in the lab before you graduate. Jenn DeBellis was highly instrumental in the design and execution of the research presented here. Her engineering prowess, critical thinking, and strong work ethic improved this project in innumerable ways. Ameet Pinto of Northeastern University offered valuable guidance on DNA sequencing and data analysis. Maria Sevillano Rivera of Northeastern University assisted in exploratory analysis of sequencing data and provided R-scripts for sequence data processing. Brian Shoener offered vital input on R coding and statistical analyses.

Additionally, I recognize the community of people who have made Urbana, IL a vibrant and enriching place to live, especially Nadine Auborg, Bao Bui, Elizabeth Chippewa, Martha Cuenca, Brighten Godfry, Kate Insolia, Janice Jayes, Susan McIntyre, Fr. Tom Royer, Stephanie Stoekel, Ian Traniello, Jane Valentine, and Carlos Villanueva. My friendship with each of you has brought much joy and light to my life.

Finally, I thank Jeff Carnes and the Daily Bread Soup Kitchen for building a place of joyful welcome for all, the Urbana Country Dancers for fostering community through movement and music, Urbana Boulders for providing a space for mountain adventure among the cornfields, and the members of the social justice reading group for insightful conversation, honesty, and novel perspectives on the world.

The work contained in this thesis was financially supported by fellowships from the UIUC Graduate College and Department of Civil and Environmental Engineering. Research funding was from NSF Grant 1-483970-251005-191100.

TABLE OF CONTENTS

Chapter 1: Introduction

Minimizing nutrient pollution to surface waters remains a persistent challenge for environmental scientists and engineers. Excess nitrogen (N) and phosphorus (P) from industrial processes, agriculture, and domestic waste (OIG US EPA, 2008) lead to eutrophication when released into surface waters, negatively impacting aquatic ecosystems, human health, recreation, and local economies (OW US EPA, 2015). Increasingly stringent N and P discharge limits have been shown to effectively prevent some of these negative water quality impacts (Conley et al., 2009), but new technologies are needed. The current limit of technology (LOT) for nutrient removal, approximately 3 mg $N \cdot L^{-1}$ and 0.1 mg $P \cdot L^{-1}$ (Bott & Parker, 2011), relies on energy-intensive and costly processes (OW US EPA, 2015). These technologies ultimately render much of the nutrients removed from the system non-bioavailable, by the biochemical conversion of inorganic N species to N_2 gas through nitrification and denitrification or chemical precipitation and sequestration of PO_4^{3-}-P into metal phosphate precipitates (Metcalf & Eddy, 2014). The Haber-Bosch process for fixing atmospheric N_2 to bioavailable NH_3 is an energy-intensive process (Cherkasov, Ibhadon, & Fitzpatrick, 2015) and mineral phosphate rock is a nonrenewable resource (Cordell & White, 2014), suggesting that a restructuring of nutrient removal technology is needed to advance sustainability goals. Recent research has begun to address this issue, seeking to reframe wastewater treatment plants (WWTPs) as water resource recovery facilities (WRRFs) (Daigger, 2017) where biosolids and dissolved nutrients are treated as resources to be recovered and reused rather than waste products for disposal (Guest et al., 2009). There is therefore a need to develop technologies which both reduce effluent nutrient loadings below the current LOT and accommodate nutrient recovery and reuse.

Algae are well equipped to address both of these challenges. The biological capabilities which cause algae to form algal blooms leading to eutrophication and hypoxia are directly related to the ability to take up nutrient to the point of environmental depletion. Research seeking to apply and contain this functionality within wastewater treatment plants has a long history. Oswald et al. (1957) first proposed the use of algae in wastewater treatment lagoons, primarily to provide oxygenation and to reduce biochemical oxygen demand. More recently, high rate algal ponds (R. J. Craggs, Davies-Colley, Tanner, & Sukias, 2003) and attached growth algal systems (Rupert J Craggs, Adey, Jessup, & Oswald, 1996) have been used for nutrient removal. However, attached growth systems have lower nutrient uptake capabilities than suspended growth systems and both suffer from inconsistent productivity and performance (Shoener, Bradley, Cusick, & Guest, 2014). Coupled with its nutrient removal capacity is the ability of

algae to synthesize storage carbohydrates and lipids which provide excess energy for nighttime nutrient uptake (Gardner-Dale, Bradley, & Guest, 2017) or provide a carbon sink for photosynthesis in cases of longer-term nutrient limitation (Dragone, Fernandes, Abreu, Vicente, & Teixeira, 2011; Markou, Angelidaki, & Georgakakis, 2012). Extensive research has gone into the field of algal biofuels, the conversion of either whole biomass or extracted carbohydrate or lipid portions into biocrude (Li et al., 2017), bioethanol (Brányiková et al., 2011), or biodiesel (Hu et al., 2008), respectively. Some have sought to identify pure algal strains or genetically engineered cultures with optimal lipid or carbohydrate composition (A. Ghosh et al., 2016; Gomaa, Al-Haj, & Abed, 2016). However, although open outdoor cultivation systems are less expensive to construct and maintain than photobioreactors (US DOE, 2016), they are not conducive to the maintenance of axenic cultures (Hoffmann, 1998; Kazamia, Riseley, Howe, & Smith, 2014). Mixed culture systems are often more robust and resilient than pure culture systems (Kazamia et al., 2014), but need to be optimized for integration into wastewater treatment. While algal functional traits are readily applicable to addressing sustainability goals and improving wastewater treatment performance, a framework for highly controlled open mixed algal systems is needed.

Environmental biotechnology harnesses knowledge of microbial metabolism to design systems which preferentially promote the proliferation of taxa with desirable functional traits (Rittmann & McCarty, 2001), and has already been applied to wastewater treatment, for example, in anaerobic digestion and biological nutrient removal (Rittmann & McCarty, 2001). Algal biotechnology could provide a middle path in algal wastewater treatment, allowing for tighter process control than open ponds without the challenges associated with pure culture growth. In a system designed based on algal biotechnology, outside strains serve merely as additional material on which the system might select (Mooij, Stouten, van Loosdrecht, & Kleerebezem, 2015), rather than as invaders which would disrupt pure culture systems. Mooij et al. (2013) proposed the ‘Survival of the Fattest’ hypothesis which outlines methods for system design favoring the survival and proliferation of algal strains with high carbohydrate or lipid storage by harnessing the connection between nutrient limitation and carbon storage in algae via the combination of daytime nutrient limitation and nighttime nutrient feeding of mixed algal cultures. Other traits such as tolerance of harsh growth conditions, biomass characteristics, and growth rate might also be selected for via cultivation under appropriate environmental conditions (Mooij, Stouten, et al., 2015). With sufficient knowledge of algal metabolism, algal biotechnology might be used to design systems which simultaneously address multiple challenges in modern wastewater treatment.

Several key factors remain to be addressed including (i) the incorporation of nutrient removal goals into the design of engineered algal systems, (ii) the need to understand similarities and differences in the response of different algal communities to a common set of growth conditions, and (iii) the need to link community structure with community function. The study outlined in the following chapters addresses each of these issues. We compare performance metrics, including nutrient dynamics, between three algal communities sourced from surface water and wastewater treatment clarifiers in three geographically distinct locations in the continental United States. We also provide in-depth analysis and comparison of community structure via 18S rRNA gene sequencing. Finally, we examine relationships between metrics of community structure and function and look for commonalities between algal communities.

Chapter 2: Background

2.1 Phototrophic Communities in Wastewater Treatment Systems. The results of several studies which have aimed to characterize algal communities present in waste treatment systems illustrate the inherent variability in such communities through space and time. Because molecular methods for describing community composition were developed only recently and are still in development for eukaryotic systems, most community composition data have been gathered via microscopy, likely resulting in a vast underreporting of rare and morphologically similar taxa. While high rate algal ponds (HRAPs) and retention ponds contain a variety of algae and zooplankton, algae tend to dominate these mixed systems (Barthel, Oliveira, & Costa, 2008; Godos, Blanco, García-Encina, Becares, & Muñoz, 2009), the composition of which can be altered by system design decisions (Park, Craggs, & Shilton, 2011, 2013). For example, in a system designed to improve settling properties of algae by recycling settled algae back into the system, the community quickly became dominated by *Pediastrun boryanum*, in contrast to a control community where *Dictyosphaerium* sp. was dominant (Park et al., 2011, 2013). *Micractinium* and *Thalassiosira* were minor taxa in these systems (Park et al., 2013).

Algal communities vary with geographic location and type of water body. For example, composition of dominant taxa differed in wastewater treatment plant (WWTP) communities sampled in Northfield, Michigan; Tampa, Florida; and Suffolk, Virginia (S. Ghosh & Love, 2011). Using two different primer sets for the *rbcL* gene, which each identified different dominant taxa in each system, there was little overlap between the locations. Michigan was dominated by *Eustigmitophyceae*, *Cylindrotheca closterium*, raphid pennate diatoms, *Chlorococcum ellipsodium*, and *Chaetomorpha linum*; Virginia by raphid pennate diatoms and *Synechocystis*, and Florida by *Nitzschia*, dinoflagellates (eg. *Pridinium*), and cyanobactieria such as *Thermosynechococcus*. However, a study using the V4 and V8-V9 regions of the 18S rRNA gene to probe community structure found many similarities between algal communities collected from wastewater treatment primary and secondary clarifiers, algaewheels, raceway ponds, and natural surface waters located in the Midwestern United States (Bradley, Pinto, & Guest, 2016). Stramenopiles, Chloroplastida, Alveolata, and unclassified Eukaryotes were common taxonomic classifications across all sample types, although with varying relative abundance. However, greater diversity and divergence between communities was evident when sequence data was classified at a finer taxonomic scale.

Individual mixed communities also vary on a longitudinal time scale in ways which can differ between parallel communities (Godos et al., 2009). Two parallel HRAPs in Valladolid, Spain

showed differing composition through time from January through September, with greatest diversity occurring at the end of the experiment, which also corresponded to the highest productivity. Numerous factors could have contributed to the observed community dynamics including variation in influent strength, weather patterns, and time from inoculation. Dominant taxa in these studies included *Chlamydomonas, Chlorella, Nitzschia, Achnanthes, Protoderma, Selenastrum, Oocystis*, and *Ankistrodesmus*.

The variation within wastewater algal communities across space and time suggests that numerous factors contribute to community dynamics. Only recently has sequencing technology afforded the ease and resolution necessary to examine these communities in-depth (Bradley et al., 2016; S. Ghosh & Love, 2011). A later section will discuss recent advances and persistent challenges in the sequencing of eukaryotic algal communities.

2.2 Nutrient Uptake Plasticity: Nitrogen and Phosphorus. Redfield proposed that algal nutrient ratios were relatively stable between species and across time (Redfield, 1958). The proposed average ratio is commonly referred to as the Redfield ratio (molar ratio of C:N:P = 106:16:1, mass ratio of C:N:P = 41.1:7.23:1) (Goldman, McCarthy, & Peavey, 1979). However, it is now well established that algal C:N:P varies significantly within and between taxa. For example, for algae grown in nutrient replete conditions, there is significant intraspecific variation in N:P between taxa (N:P of 5 to 19 $mol \cdot mol^{-1}$, 2.26 to 8.59 $g \cdot g^{-1}$), with most values falling below the Redfield ratio (Geider & LaRoche, 2002). Under nutrient deplete conditions, however, the N:P range is much wider, from <5 to >100 $mol \cdot mol^{-1}$ (<2.26 to >45.2 $g \cdot g^{-1}$) (Geider & LaRoche, 2002), indicating that N and P dynamics are highly plastic. This plasticity is driven by a combination of nutrient supply ratio and growth rate (Klausmeier, Litchman, Daufresne, & Levin, 2008) with slower growth rates being associated with higher carbon to nutrient ratios (e.g., under N-limited conditions, C:N increased from 7.1-20 $mol \cdot mol^{-1}$ (2.75-7.75 $g \cdot g^{-1}$) with decreasing growth rate from 90% to 10% of maximum) (Geider & LaRoche, 2002) and nutrient limitation associated with decreased proportion of the limiting nutrient in comparison to other components (e.g., N:P increased from 15 to 115 $mol \cdot mol^{-1}$ (6.78 to 51.99 $g \cdot g^{-1}$) under P-limited conditions as growth rate decreased from 90% to 10% of maximum) (Geider & LaRoche, 2002). For a given nutrient supply ratio, biomass N:P increases linearly with growth rate, although the magnitude of this relationship varies by species (Gardner-Dale et al., 2017). If the nutrient supply ratio falls within the N:P plasticity for a given culture, it will transition from being N-limited at higher N:P and growth rate to being P-limited at lower N:P and lower growth rate, passing through a point of N and P co-limitation, where biomass and media N:P are equal (Gardner-

Dale et al., 2017). This N:P ratio, known as the critical ratio, varies between approximately 20 and 50 mol·mol^{-1} (9 and 22.6 g·g^{-1}) (Geider & LaRoche, 2002) and represents a point at which both N and P would need to increase in order to achieve an increase in growth rate. Under conditions of extremely scarce N and P, biomass N:P tends to match that of the environment (Geider & LaRoche, 2002). To understand these relationships from a functional perspective, it is helpful to understand how C, N, and P are assimilated into cell biomass.

Cellular carbon, nitrogen, and phosphorus are partitioned between structural components, nutrient reserves, and energy stores (Geider & LaRoche, 2002). Allocation of C, N, and P is related to the cellular processes required under a given nutritional status and growth rate. Nutrient-related cell processes and associated structures may be divided into the categories of uptake (proteins and chloroplasts) and assimilation (ribosomes) (Klausmeier et al., 2008). In nutrient replete conditions when uptake is most favorable, more energy and resources are allocated to maximizing assimilation and growth rate. However, under nutrient limited conditions, competition for scarce resources is prioritized (Klausmeier et al., 2008) and priority is given to nutrient uptake. Ågren (2004) determined that N:C increases linearly while P:C increases quadratically with growth rate, relationships which may be explained by the structural role of N-rich protein for the construction of new cells and of P-rich ribosomal RNA required for protein synthesis. While there are physiological constraints on cellular composition, relative proportions of major cellular components – DNA, RNA, lipids, carbohydrates, proteins, pigments, and other small molecules – still vary widely in relative proportion of cellular dry weight (Geider & LaRoche, 2002). Range estimates for N allocation include: free amino acids and proteins (65-85%), RNA (1.3-13%), DNA (0.4-5%), and Chlorophyll (0.2-3%) (Lourenço, Barbarino, Marquez, & Aidar, 1998). For P allocation, these estimates are less well-established, but include RNA (30-100%) and phospholipid (<5%) (Rhee, 1978). When only one nutrient is limiting, luxury uptake and storage may account for a significant proportion of intracellular N (6-36%) and P (>40%) stored as NO_3^--N and polyphosphate, respectively (Lourenço et al., 1998; Rhee, 1978). Storage carbohydrates and neutral lipids may each account for 10-50% of cell biomass, but contain no N or P and thus do not impact N:P. The impact of N and P availability on carbon storage as carbohydrate and lipid fractions will be discussed in the following section.

2.3 Algal Carbon Storage. The bulk of algal biomass consists of proteins, carbohydrates, and lipids. While proteins play a dominant structural role, carbohydrates and lipids serve both structural and storage functions. Storage carbohydrates and lipids consist of carbon, hydrogen, and oxygen and thus function as a carbon sink for photosynthesis in times of nutrient shortage.

The factors influencing the synthesis and function of carbohydrates and lipids are outlined below.

Algal carbohydrates are comprised of a number of monomers including xylose, mannose, glucose, galactose, and rhamnose, which vary in relative abundance by species (Markou et al., 2012). Algal cell walls may contain cellulose or hemicellulose as structural components (González-Fernández & Ballesteros, 2012). Under optimal growth conditions, algae shunt some fixed carbon into starch during the day to provide an energy source for respiration during the night (Gardner-Dale et al., 2017; González-Fernández & Ballesteros, 2012). Storage carbohydrates are also synthesized in response to environmental stressors including nutrient depletion (N, P, S, Fe), low ambient inorganic carbon, and high salinity (Markou et al., 2012). In one study, *Chlorella vulgaris* starch content increased from 5% to 60%, 55%, and 35% under S-, P-, and N-limitation, respectively (Brányiková et al., 2011). Elsewhere, carbohydrate content of three algal taxa increased from 8-17% up to 70% under N-starvation (Markou et al., 2012).

Lipids synthesized by algae include neutral and polar lipids, wax esters, sterols, hydrocarbons, prenyl derivatives, and phytylated pyrrole derivatives (e.g., chlorophyll) (Hu et al., 2008). The synthesis of fatty acids (FAs), the building blocks of lipids, begins in the chloroplast (Hu et al., 2008). Carbon fixed by photosynthesis is shunted to the lipid pathway via the synthesis of malonyl CoA from acetyl CoA (Hu et al., 2008). Under optimal growth conditions, most FAs are converted to membrane lipids which constitute between 5% and 20% of algal dry weight (Hu et al., 2008). Most membrane lipids contain between 10 and 20 carbon atoms and are polyunsaturated (Hu et al., 2008). However, as with carbohydrates, environmental stress may shift algal metabolism to favor the synthesis and storage of neutral lipids, mostly triacylglycerols (TAGs) (Hu et al., 2008). While most algae contain TAG stores even under optimal growth conditions (e.g., between approximately 5% and 85% with a mean of 25.5% for green algae and a range of approximately 15% to 40% with a mean of 22.7% for diatoms), neutral lipid synthesis may be greatly enhanced (e.g., between approximately 15% and 90% with a mean of 45.7% for green algae and between approximately 20% and 80% with a mean of 37.8% in diatoms) in some algal taxa under conditions of environmental stress (Hu et al., 2008). As with carbohydrates, N, P, and SO_4^{2-} depletion, as well as silicon limitation in diatoms, has been shown to lead to lipid accumulation (Guschina & Harwood, 2006; Hu et al., 2008; Markou et al., 2012). Additionally, lower temperatures may induce more unsaturated FAs, lower light may lead to polar lipid formation for chloroplast synthesis, and higher light may cause more TAG production. Increased culture age is also associated with increased TAG production (Hu et al.,

2008). Extant lipid content under ideal growth conditions and magnitude of response to individual stressors varies widely in magnitude among algae in a manner which is largely species or strain specific, and not generally predictable based on higher levels of taxonomic classification (Hu et al., 2008). Similarly, FA characteristics and the impacts of growth conditions on FA composition vary widely (Hu et al., 2008). Detailed descriptions of lipid synthesis pathways are included in Hu et al. (2008), Harwood and Jones (1989), and Guschina and Harwood (2006).

Environmental stress, including nutrient limitation, induces algal cultures to accumulate both starch and lipids. However, the relative potential for carbohydrate and lipid accumulation varies greatly by strain. To better understand, and potentially engineer, mixed algal systems, a more developed understanding of the biochemical and genetic controls which drive carbon storage is needed.

2.4 DNA Sequencing Technology. Due to the wide variation in functional traits among algae described above coupled with the need for tight control of system performance, the ability to link the structure and function of algal communities would be beneficial to the design and operation of algal systems. While the algal community composition of wastewater systems has historically been documented via microscopy (eg., Barthel et al., 2008), this is a time-consuming process which requires highly trained personnel and lacks the ability to distinguish between morphologically similar taxa (Eland, Davenport, & Mota, 2012). Recent advances in sequencing technology, however, enable collection of unprecedented quantities of sequencing data and generate new insights into the structure of microbial communities. However, much of this work has historically focused on bacterial rather than eukaryotic communities and careful data processing is required for accurate interpretation of the results of sequencing studies. The following discussion details the current state of sequencing technology as it applies to eukaryotic communities and highlights persistent challenges in the field.

Sequencing technology has improved rapidly over the past several decades. Initially, Sanger sequencing required the assembly of clone libraries in a labor-intensive process which allowed for little depth of sampling and was subject to significant bias (Schloss & Westcott, 2011). More recently, 454 pyrosequencing provided a high-throughput platform which allowed for the generation of large quantities of sequencing data with comparatively little labor input (Ronaghi, 2001). Most recently, Illumina MiSeq high-throughput sequencing technology provides higher throughput and more rapid data generation (Lin Liu et al., 2012). As technology has evolved, so has the way in which target genes for sequencing are selected.

All of the technologies described above rely on the polymerase chain reaction (PCR) to amplify DNA from a biological sample before sequencing. For sequencing of prokaryotic communities, well-developed sequencing methods for the 16S rRNA gene (Gu, Nerenberg, Sturm, Chul, & Goel, 2010) have, for example, been applied to the study of engineered systems such as anaerobic digesters (e.g., Vanwonterghem, Jensen, Ho, Batstone, & Tyson, 2014) and drinking water distribution systems (e.g., Pinto, Schroeder, Lunn, Sloan, & Raskin, 2014). However, methods for eukaryotic systems are less well established and several factors must be considered in the selection of a region of the genome for PCR amplification. A target DNA region should be selected based on the intended scope of the analysis. Not only is it important that the gene be ubiquitous in all organisms within the group of interest but that the rate of mutation allow for the desired level of resolution. The trade-off between taxonomic resolution and universal application is illustrated in a recent study of the Bacillariophyta (diatoms) by Guo et al. (2015) in which genes for the cytochrome c oxidase 1 (COI), part of the *rrn* operon (5.8S + internally transcribed spacer (ITS)), 18S rRNA gene, and *rbcL* gene coding for RuBisCO were compared. They found that COI and the *rrn* operon allowed for accurate resolution of the taxonomy of closely related taxa but could not distinguish taxonomic relationships across all Bacillariophyta. The 18S gene, conversely, had a lower mutation rate resulting in errors in establishing relationships between closely related taxa but allowing for its broad application across more distantly related taxa. The 18S gene has the additional desirable quality of containing numerous alternating conserved and variable regions. While Sanger sequencing allowed for the sequencing of long DNA template regions (e.g., 1800 bp , 560 bp, and 950 bp of the 18S rRNA gene by (Bazin, Jouenne, Deton-Cabanillas, Pérez-Ruzafa, & Véron, 2014; Eland et al., 2012; Viprey, Guillou, Ferréol, & Vaulot, 2008), MiSeq technology is limited to segments of 250-300 reads in length, depending on the chemistry used (https://www.illumina.com). Numerous variable regions of the 18S rRNA gene have been sequenced including V1-V2 (Mohrbeck, Raupach, Martínez Arbizu, Knebelsberger, & Laakmann, 2015), V3 (Medinger et al., 2010), V4 + V9 (Lohan, Fleischer, Carney, Holzer, & Ruiz, 2016; Stoeck et al., 2010), V4 + V8-V9 (Bradley et al., 2016), and V9 (Vargas et al., 2015). In cases where V4 and V8-V9 or V9 were sequenced together, V8-V9 or V9 was found to achieve better representation across widely differing taxa while V4 achieved better resolution of closely related strains (Bradley et al., 2016; Stoeck et al., 2010).

Even with a well-selected target region and primers, sequencing is still subject to PCR bias and sequencing error which negatively impact data quality. Bias associated with PCR begins with DNA extraction, where it is possible for some methods to inadequately lyse tough cell walls

(Eland et al., 2012; Medinger et al., 2010). This may be minimized by selecting a DNA extraction method adequately tested for the type of sample being analyzed. Primer bias occurs when primers have unequal affinity for the suite of taxa being analyzed. Bradley et al. (Bradley et al., 2016) found that a universal primer designed to target the V4 region of the 18S rRNA gene had uneven coverage of a mock community designed to encompass a broad diversity of freshwater and marine algal taxa. To improve primer coverage, the primer was modified by removing a degeneracy at the 3' end of the reverse primer which allowed it to bind to both guanine (G) and adenine (A). Others have also found that degenerate bases should be avoided in primers since guanine/cytosine (G/C) and adenine/thymine (A/T) base pairs differ in bond strength (Polz & Cavanaugh, 1998). The modified primer contained an exact match to all mock community members and provided a more accurate representation of the mock community (Bradley et al., 2016). Base composition of the target region also influences relative PCR efficiency, with higher GC content diminishing PCR efficiency due to greater bond strength and correspondingly higher melting temperature (Benita, Oosting, Lok, Wise, & Humphery-Smith, 2003). Bradley et al. (2016) found that high GC content was likely the cause of poor coverage of haptophyte algae (*Prymnesium parvum* and *Isochrysis galbana*) using modified V4 primers. A proposed solution is to optimize thermocycler settings for different GC content thresholds and to run mixed samples under the full range of settings to ensure adequate amplification of taxa encompassing the full range of GC content (Benita et al., 2003). PCR selection – the overamplification of DNA from more abundant sequences such that lower-abundance sequences are crowded out – may be minimized via the combination of higher concentrations of template DNA with decreased number of PCR cycles (Polz & Cavanaugh, 1998). PCR drift – the inadequate representation of a sample based on non-representative sampling of the template stock – may be minimized by pooling replicate PCR amplifications of a given stock prior to sequencing (Polz & Cavanaugh, 1998). It is possible to mitigate much PCR and primer bias using the techniques listed above, although specific concerns will vary with experimental conditions.

In addition to influencing PCR amplification efficiency, primers and primer design also play a key role in sequencing. Since individual amplified PCR samples are pooled before MiSeq sequencing, it is important to be able to link each sequence to its original sample. This is often accomplished using dual index barcodes, which are unique base sequences added to the 5' end of forward and reverse primers such that amplified sequences may be identified by the unique combination of forward and reverse primer barcodes (Huber et al., 2007). Additionally, in low-diversity communities, highly homogenous base calls in the initial stages of sequencing on the

Miseq can cause instrument calibration errors which result in poor quality or absent data (Bradley et al., 2016). This problem may be mitigated by adding variable length linker sequences to primers such that the signal received by the system in the initial stages is less homogenous, resulting in better system calibration and higher data quality (Bradley et al., 2016).

High throughput sequencing is designed for data quantity over quality (Schloss & Westcott, 2011) and sequencing errors may impair accurate assessment of community composition and inflate diversity metrics. Improved data processing can minimize these errors and improve the accuracy of results (Schloss & Westcott, 2011). Types of sequencing errors may include ambiguous base calls, base substitution errors, abnormally long homopolymer chains, and chimeras. Chimeras refer to sequences which may be formed and amplified in PCR when a sequence extends only partially and then re-anneals with a different sequence during the subsequent PCR cycle, resulting in amplified sequences which are the result of multiple parent strands. These types of sequencing errors may largely be removed via a well-designed data post-processing pipeline.

Aspects of the design of such a pipeline are described by Schloss and Westcott (2011) and integrated into a formalized pipeline for use with the Mothur software package by Kozich et al. (2013). Highlights of this pipeline, which can reduce sequence error by an order of magnitude (Schloss & Westcott, 2011), are outlined below. They suggest removing sequences containing ambiguous base calls, >1 barcode mismatch and >2 primer mismatches along with sequences containing homopolymer sequences exceeding 8 bases. They apply quality trimming using a sliding window approach to truncate sequences if they fail to meet a given quality threshold. Additionally, they define a minimum overlap requirement for forward and reverse reads, trim stitched sequences to start and end at the same position, and use denoising algorithms to precluster data based on maximum difference thresholds. While these steps successfully removed >75% of chimeras (Schloss & Westcott, 2011), they identify UCHIME as the optimal existing algorithm for identification and removal of remaining chimeras. UCHIME organizes sequences in order of abundance and divides each less abundant sequence into chunks along its length, comparing each chunk with more abundant sequences. Sequences are tagged as chimeras when they are comprised of chunks matching with two or more parent segments. Despite these measures, some erroneous sequences will inevitably remain in the dataset but their impact may be minimized via use of optimized clustering techniques to reduce spurious OTUs and taxonomic classifications (Schloss & Westcott, 2011).

Numerous clustering algorithms exist which cluster sequences into OTUs based on a user-defined minimum similarity threshold. The current default method in Mothur is OptiClust (Westcott & Schloss, 2017), which outperforms other clustering methods by maximizing clustering quality and minimizing time and RAM required to complete the analysis. OptiClust maximizes the Mathews correlation coefficient (MCC, a metric of assignment quality) in real time by moving individual sequences from one cluster to another only if doing so increases the MCC. While a 97% sequence identity clustering threshold for OTU formation is commonly used in the literature (Huber et al., 2007; Kunin, Engelbrektson, Ochman, & Hugenholtz, 2010), some have suggested using a mock community to define optimal thresholds for OTU formation and other relevant metrics within each sequencing run such as quality threshold and sequence overlap (Bradley et al., 2016). Additionally, Schloss and Westcott (2011) suggest using mock communities to calculate rates of chimera formation, sequencing error, and drift, allowing for better tracking and comparison of error between runs.

Despite these quality control techniques, a persistent issue with MiSeq sequencing is significant variation between technical replicates (Wen et al., 2017). Improved depth of sequencing as well as pooling of DNA extracts and PCR products can help minimize this effect. They also suggest removing OTUs containing only a single sequence and OTUs which are present in only a single technical replicate, as these are more likely to be errors as opposed to rare OTUs (Wen et al., 2017).

A persistent challenge in the use of 18S sequencing data to describe eukaryotic microbial communities is the wide range in 18S rRNA gene copy number between taxa, ranging from <10 to >10,000 (Vargas et al., 2015; Zhu, Massana, Not, Marie, & Vaulot, 2005). While Viprey et al. (2008) confirmed that 18S gene copies within an organism are identical and will not inflate diversity dynamics, copy number still poses problems in describing community structure based on read counts. Vargas et al. (2015) determined that copy number is a better indicator of relative biovolume or biomass than of relative cell abundance. Zhu et al. (2005) and Vargas et al. (2015) attempted to fit a linear model to the log-log relationship of cell size and copy number, although this model has a wide confidence interval, especially at large cell sizes due to the logarithmic scale. The copy number problem needs to be addressed in order to more accurately describe eukaryotic community composition based on cell counts rather than biomass-based metrics. It might be worthwhile to consider other genes which have more consistent copy number across eukaryotic taxa. Additionally, a relevant question for environmental engineers might be whether the use of community composition based on cell counts to describe eukaryotic

communities is the most relevant metric. In community and population ecology, the individual organism is an important functional unit. However, from a standpoint of system function and biomass production, the cells which represent the greatest quantity of biomass are likely to drive function (biomass composition and nutrient uptake), suggesting that relative biomass contribution may be a sufficiently relevant metric for environmental engineers.

Rapidly advancing high-throughput sequencing technology is a valuable tool for microbial ecology and has the potential to greatly enhance our understanding of microbial communities in natural and engineered systems. While this technology is better-developed for prokaryotes, significant improvements have recently been made in the sequencing of eukaryotic communities. New high-throughput sequencing technologies make it easier and faster to generate large quantities of sequence data while advances in primer design, template region selection, and data processing algorithms promote continual improvements in data quality. Continuing research will be required to resolve ongoing challenges in PCR bias, data processing, and the copy number problem which currently impair the capacity to quantitatively describe eukaryotic community composition.

Chapter 3: Impact of Daytime Nitrogen Limitation on Structure and Function of Mixed Microalgal Communities Derived from Three Geographically Distinct Sources

**The following section contains a manuscript that was in preparation at the time of the submittal of this thesis. It is included in its entirety below.*

Authors

Anna C. Fedders[1], Jennifer L. DeBellis[1], Ian M. Bradley[1], Maria Catalina Sevillano Rivera[2], Ameet J. Pinto[2], Jeremy S. Guest[1,*]

Author Affiliations

[1] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA

[2] Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, 02115, USA

*Corresponding Author: jsguest@illinois.edu

Keywords: algae, phytoplankton, wastewater treatment, mixed culture, photobioreactor

Abstract

Increased nutrient pollution in recent decades has been a driver of persistent eutrophication and hypoxia in aquatic systems, to the detriment of ecosystem function, human health, and local economies. These problems have been successfully addressed via regulations mandating reductions in effluent nitrogen (N) and phosphorus (P), including from wastewater treatment plants (WWTPs). However, current means of N and P removal in WWTPs render removed nutrients non-bioavailable and are incompatible with sustainability goals aiming for nutrient recovery and reuse combined with reduced energetic requirements. Eukaryotic microalgae are uniquely suited to address this challenge due to their ability to remove dissolved N and P to limiting levels and to accumulate carbohydrate and lipid stores for potential downstream conversion to biofuels. Research on algal use within these contexts has historically focused on minimally engineered open outdoor systems for nutrient removal or axenic cultures for biofuel production. Algal biotechnology, however, provides an alternative path by which mixed culture cultivation systems are designed to allow for the preferential proliferation of algal taxa with desirable functional traits. In this study, we tested the ability of three algal communities sourced from geographically distinct latitudes within the continental United States (Tampa, Florida; Durham, North Carolina; Urbana, Illinois) to provide consistent nutrient removal and carbon storage under identical growth conditions. Cultures were grown with an 8-day solids residence time (SRT) and 14:10 light:dark cycle. Media had a 2:1 N:P ratio and cultures were operated under nighttime N-feeding and daytime N-limitation for 10.25 SRT, followed by a 10-day batch experiment to measure kinetic storage parameters. Media and biomass properties were measured for the duration of the study and community composition was determined for the 10.25 SRT experiment via 18S rRNA gene sequencing of the V4 and V8-V9 regions using Illumina MiSeq technology. Function and structure followed complementary patterns with similar transition points for the duration of the long-term experiment, at the end of which, algae exhibited nutrient dynamics (biomass N:P, biomass-specific PO_4^{3-}-P uptake) and storage characteristics (carbohydrate and lipid) which were similar between reactors, despite persistent differences in community structure between reactors. Intrinsic storage capacity for both lipids and carbohydrates, however, differed widely between reactors. Our results suggest that algal communities with persistently different structures can achieve consistent performance with respect to nutrient dynamics and extant carbon accumulation but still differ widely with regard to intrinsic storage parameters.

3.1. Introduction. The rapid increase in nitrogen (N) and phosphorus (P) loadings to surface waters over the past several decades (OW US EPA, 2011) has been a primary driver of widespread toxic algal blooms, eutrophication, and hypoxia (Conley et al., 2009; Paerl, Valdes, Joyner, Piehler, & Lebo, 2004; Rabalais, Turner, & Wiseman, 2002) to the detriment of aquatic ecosystems, human health, and local economies (OW US EPA, 2015). However, improved management strategies, including increasingly strict point-source discharge limits for for N and P, (Clark et al., 2010) can help mitigate these negative impacts (Hendriks & Langeveld, 2017; Paerl et al., 2004; Ruhl & Rybicki, 2010). Technological advancements which, in recent years have brought the limit of technology (LOT) for nutrient removal to approximately 3 mg $N \cdot L^{-1}$ and 0.1 mg $P \cdot L^{-1}$, (Bott & Parker, 2011; OW US EPA, 2011) have also been costly and energy-intensive (Clark et al., 2010; OIG US EPA, 2008). Thus, an important engineering challenge in modern wastewater treatment is to advance goals for local and regional water quality while also navigating tradeoffs for broader initiatives for utility and industry environmental, economic, and social sustainability (Guest et al., 2009).

Phototrophic bioprocesses are a promising means of nutrient removal from waste streams (Shoener et al., 2014) due to the ability of microalgae to assimilate N and P to the point of depletion (Gardner-Dale et al., 2017; Reynolds & Likens, 2009). Integration of algae into wastewater treatment has historically focused on open systems such as facultative lagoons (US EPA, 2002) and, more recently, high rate algal ponds (HRAPs) (Shoener et al., 2014) which often suffer from low productivity (Chisti, 2007) and inconsistent nutrient removal (R. J. Craggs, Heubeck, Lundquist, & Benemann, 2011). In a comparison of lagoons, HRAPs, and more highly engineered cultivation systems, photobioreactors (PBRs) yielded the greatest nutrient removal (Shoener et al., 2014). While PBRs also required the greatest energetic input of the compared systems, there is potential for this to be offset in part via the conversion of algal biomass to biofuels (Shoener et al., 2014). Additionally, individual algal strains have been studied for nutrient uptake and carbon storage capabilities (Sydney et al., 2011) but pure cultures are susceptible to contamination and culture crash in open systems with nonsterile media (Davis, Markham, Kinchin, Grundl, & Tan, 2016) making them impractical for wastewater treatment. Environmental biotechnology, however, has the flexibility to optimize the function of any growth system by applying knowledge of algal metabolism to process design which favors the preferential proliferation of (i.e., selects for) algae exhibiting the optimal expression of ubiquitous algal traits such as N and P uptake and carbon storage (Guest, van Loosdrecht, Skerlos, & Love, 2013; Mooij et al., 2013). Mooij et al. (2013, 2015, 2016) tested such a system in which daytime N-limitation was used to select for algal strains exhibiting elevated carbohydrate

content with a goal of producing biofuel feedstock, but without a focus on tracking nutrient dynamics or directly comparing algal communities derived from different sources. Increased understanding of these phenomena will be a critical step towards the goal of using algal biotechnology to model and design treatment systems for reliable performance irrespective of geographic location.

The objective of this work was to compare community structure and function between algal cultures derived from geographically diverse locations and maintained under a nutrient regime designed to select for carbohydrate accumulating strains (Mooij et al., 2013). System function was described in terms of nutrient dynamics and algal biomass composition, while 18S rRNA sequence data provided structural metrics. Three sequencing batch reactors (SBRs) were inoculated with naturally occurring algae sourced from surface water and/or wastewater treatment plant (WWTP) clarifiers from one of three different geographic locations within the United States (Tampa, Florida; Durham, North Carolina; Urbana, Illinois). In doing so, the aim was to incorporate the spatial heterogeneity of ambient cultures which would naturally be available to seed open algal cultivation systems in different locations. Reactors were operated under daytime N-starvation and nighttime N-feeding, after which a kinetic batch experiment was conducted to characterize intrinsic kinetic parameters and carbon storage capabilities of each culture. Results of this study will inform system design which aims to achieve consistent performance irrespective of geographic location.

3.2. Methods.

3.2.1 Photobioreactors (PBRs). Flat panel PBRs were constructed as described by Gardner-Dale and colleagues (2017) (Figure A.1 in the Appendix) with the goal of maintaining well-defined conditions within the reactor while also preventing cross-contamination. Briefly, PBRs had a total volume of 4.7 L but were operated at a working volume of 4.0 L. Lighting was provided from one side with a panel of alternating strips of red (630 nm) and blue (460 nm) light emitting diodes (Blaze™ 12V LED Tape Light, Elemental LED) with light intensity controlled by a microcontroller (Red Board, Sparkfun Electronics). Mixing was achieved via continuous aeration at approximately 0.1 $L_{air} \cdot L_{reactor}^{-1} \cdot min^{-1}$ with humidified air (Guest et al., 2013). pH was maintained between 7.00 and 7.75 by adding 100% CO_2 to the aeration mix when the pH reached 7.75 (Eutech Instruments, pH 190 Series). Throughout all experiments, attached growth was aseptically resuspended using a magnetic stirbar inside each reactor manipulated with a strong magnet from outside the reactor.

3.2.2 Inocula Collection, Preparation, and Acclimation. Inocula were sourced from natural surface waters (SWs) as well as primary and secondary clarifier weirs at water resource recovery facilities (WRRFs; a.k.a., wastewater treatment plants) in three geographically distinct locations in the continental United States in late October, 2015: Tampa, Florida (FL; 28° N), ; Durham, North Carolina (NC; 36° N); and Urbana, Illinois (IL; 40° N)). Cell counts were performed on each sub-sample of inoculum, including: three SWs and one primary clarifier for IL; five SWs for NC; four SWs and one secondary clarifier for FL. A mixed inoculum was prepared for each location (FL, NC, IL) by combining sub-samples in equal cell concentrations. Finally, each mixed inoculum was added to PBRs to a final cell density of 110 cells mL^{-1}. The low cell density was necessary to accommodate natural stream samples from NC, which had very low biomass concentrations. The microbial communities were acclimated to laboratory conditions over a two-week period (Section S1, SI), after which long-term operation and sampling began.

3.2.3 Long-Term Operation and Short-Term Kinetic Assay. For long-term operation, all three PBRs (FL, NC, IL) were operated as sequencing batch reactors with a hydraulic retention time (HRT) and solids residence time (SRT) of 8 days and a 24-hour cycle. Specifically, 0.5 L of reactor volume was wasted at the start of each dark phase (i.e., immediately after simulated sunset) and was immediately replaced with the addition of 0.5 L of new medium. A sinusoidal light curve 14:10 (light:dark) was imposed with maximum light intensity of 190 $\mu E \cdot m^{-2} \cdot s^{-1}$, reactors were sampled daily for SRT 1 and 2, every two days for SRT 3 and 4, and roughly once per SRT for the remainder of the experiment. After 82 days (> 10 SRTs) of operation, a kinetic experiment was run to determine the intrinsic growth and carbon storage capabilities of each microbial community (Guest et al., 2013). At the end of the light period, medium was spiked into each reactor to achieve a final concentration of 10 $mg\text{-}N \cdot L^{-1}$ (all other media components were scaled to maintain previous operating concentrations). Light intensity increased at a rate consistent with the sinusoidal curve and, after plateauing at 190 $\mu E \cdot m^{-2} \cdot s^{-1}$, was maintained at 190 $\mu E \cdot m^{-2} \cdot s^{-1}$ for the duration of the kinetic study (244 hours).

3.2.4 Sample Collection and Storage. During long-term operation, biomass (for solids analysis, biochemical and elemental composition, and DNA extraction) was collected at the end of the light phase (immediately before simulated sunset) and aqueous samples were collected both at the end of the light phase and the end of the dark phase. Settled and fixed growth algae were resuspended before effluent was sampled into a clean beaker from a port installed at the base of each. Algae were kept in suspension using a magnetic stirbar and stirplate while being

subsampled for subsequent analyses. Samples for solids analysis were analyzed immediately (i.e., no storage). For biochemical composition and aqueous analyses, biomass was pelleted via centrifugation at 10,000 xg for 10 min, after which supernatant was filtered through pre-rinsed 0.22 µm nitrocellulose filters and stored at -20°C until analysis. Biomass pellets were stored at -20°C until being lyophilized, ground with mortar and pestle, and stored under desiccation at room temperature. For DNA analysis, 4 mL samples of algal suspension were centrifuged at 10,000 xg for 10 min, supernatant was decanted, and pellets were stored at -80°C until DNA extraction.

3.2.5 Solids Analysis. Total suspended solids (TSS) concentration was measured in triplicate for each reactor by filtering 5 mL of well-mixed reactor effluent through a pre-rinsed, ashed, pre-weighed glass-fiber filter (0.7 µm pore size, Fisher Scientific 09-804-142H). Filters were heated at 105°C for one hour before being cooled in a desiccator for 30 min and re-weighed Pruvost, Van Vooren, Cogne, & Legrand, 2009). Volatile suspended solids (VSS) concentration was determined by combusting dried filters at 550°C for 20 min, cooling and reweighing as above.

3.2.6 Aqueous Analyses. Soluble ammonium (NH_4^+) and orthophosphate (PO_4^{3-}) were determined in triplicate using the phenate method (4500-NH3, F) (APHA et al., 2012) and ascorbic acid method (4500-P, E), (APHA et al., 2012) respectively, each modified for a microplate (Gardner-Dale et al., 2017). Method detection limits have been previously determined to be 0.012 mg NH_4^+-N·L^{-1} and 00.011 mg PO_4^{3-}-P·L^{-1} (Gardner-Dale et al., 2017). Nitrate (NO_3^-) and nitrite (NO_2^-) concentrations were determined via ion chromatography (Dionex ICS 1000) and anion exchange column (Thermo Scientific IonPac IS14A).

3.2.7 Biochemical and Elemental Characterization of Biomass. Lyophilized biomass samples were analyzed for carbon, hydrogen, and nitrogen (Perkin Elmer 2400 Series II CHNS/O Elemental Analyzer) and total phosphorus (Perkin Elmer SCIEX ELAN DRC-E ICP-MS) by the Microanalysis Laboratory at the University of Illinois at Urbana-Champaign (UIUC) School of Chemical Sciences. Biomass protein content was estimated by multiplying the elemental N percentage by a conversion factor of 6.2 representing the ratio of algal protein to N content (Becker, 1994). Triplicate analyses of biomass carbohydrate content were conducted via two-step acid digestion followed by colorimetric analysis with 3-methyl-2-benzothiazolinone hydrazine (MBTH) (van Wychen & Laurens, 2015), modified for microplate reader (Gardner-Dale et al., 2017). Total biomass lipid content was determined via the Folch method (Folch, Lees, & Stanley, 1957) as modified by Axelsson and Gentili.(2014). Datasets were compared

between reactors using two-tailed ANOVAs followed by Tukey's tests with $\alpha = 0.05$, unless otherwise specified.

3.2.8 DNA Extraction, PCR, and Sequencing. DNA extraction, PCR, and product purification were conducted as previously described (Bradley et al., 2016). Briefly, DNA was extracted from ten samples spaced throughout long-term experiment for each of the three reactors using the MP BIO FastDNA SPIN extraction for soil (MP Biomedicals, Santa Ana, CA) and stored at -20 ^{o}C.

DNA extracts were amplified in triplicate via PCR using the KAPA HiFi Hotstart PCR kit (Kapa Biosystems, Wilmington, MA) using primers for the V4 and V8-V9 regions of the 18S rRNA gene (V4: forward (Reuk454FWD1) = CCGCASCYGCGGTAATTCC, reverse (V4r) = ACTTTCGTTCTTGAT, V8-V9: forward = (V8f) ATAACAGGTCTGTGATGCCCT, reverse = (1510r) CCTTCYGCAGGTTCACCTAC), reagent concentrations, and thermocycler settings, as described in Bradley et al. (2016). Mock community samples corresponding to MC4 (Bradley et al., 2016) and negative controls containing no added DNA were run in triplicate alongside samples. Gel electrophoresis was used to confirm length of PCR products and excised bands were purified using a QIAquick gel purification kit (Qiagen, Valencia, CA). The DNA concentration of each sample was measured in triplicate using Qubit 2.0 (Invitrogen). Sequencing was performed by the Roy J. Carver Biotechnology Center at UIUC using Illlumina MiSeq with v2 chemistry and 2x250 paired reads.

3.2.9 Sequence Data Processing and Analysis. Bcl2fastq v1.8.4 Conversion Software (Illumina) was used to demultiplex raw sequence data. Data was processed through Casava 1.8 (Illumina). Sickle was used to remove all bases with a phred score of less than 25 and to define a minimum read overlap requirement of 70 bp, these values determined via mock community analysis (Section S3, Figure S2). Mothur v.3 was used for sequence processing as described by Kozich et al. (Miseq SOP at http://www.mothur.org/wiki/MiSeq_SOP) using default settings unless otherwise specified (Kozich et al., 2013). Reads containing ambiguous bases were removed following contig formation. Sequences were trimmed using screen.seqs to remove sequences not starting before position 4359 and ending before position 8460. Reads were aligned with SILVA v123 and the alignment trimmed using vertical = T and trump=. commands. UChime was used to detect and remove chimeras. Singletons were removed and only remaining sequences were used in subsequent analyses. OTUs were clustered at 97% read similarity. OTUs which were present in only one of three technical replicates were omitted from further analysis using the vegan package in R (Oksanen et al., 2017). Silva v123 from Mothur

was used to assign taxonomy to the OTUs. Alpha diversity, as described by observed OTUs, Chao1 index, inverse Simpson index, and nonparametric Shannon index, were determined using the summary.single command with a subsample size equal to the smallest sample size (3,478 (V4) and 2,062 (V8-V9)). Beta diversity, as described by Bray-Curtis distance, was calculated using the summary.shared command, using the same subsample sizes listed above. The dist.shared command was used to produce Bray-Curtis distance matrices. NMDS plots were generated in Q-mode using raw data and Bray-Curtis dissimilarity with the vegan package in R (Oksanen et al., 2017). Sample richness was computed with the rarefy function from the vegan package implemented in R, based on the minimum read number on a per sample basis for both V4 and V8-9 regions of the 18S rRNA gene.

3.3. Results and Discussion. To facilitate discussion, the long-term experimental period was divided into three phases of operation based on performance: (Phase I, rapid change) SRT0 to SRT2; (Phase II, moderate change) SRT2.0 to SRT7.75; (Phase III, stable performance) SRT7.75 to SRT10.25.

3.3.1 Nutrient Recovery.

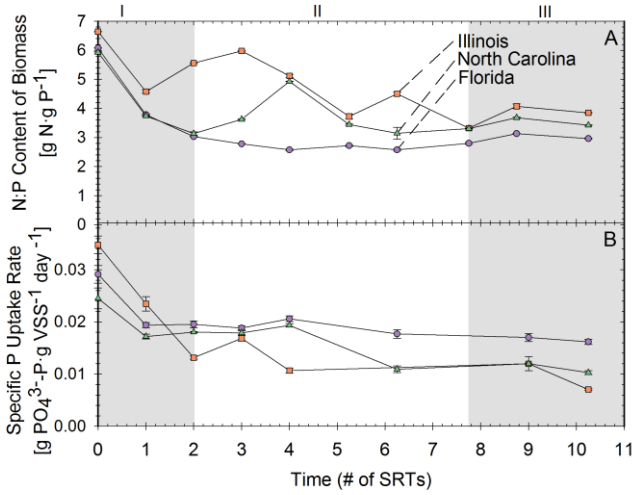

Figure 3.1: Performance of parallel PBRs inoculated with environmental samples from Florida (purple circles), North Carolina (green triangles), and Illinois (orange squares). (A) Biomass N:P content (mass/mass) and (B) observed specific PO_4^{3-} uptake rate (g PO_4^{3-}-P·g VSS^{-1}·day^{-1}). One SRT is equal to 8 days. Biomass N:P content varied across the 82 days of the study, but all three reactors tended toward a similar steady-state N:P. Phosphate recovery performance demonstrated similar trends in all three reactors, with values decreasing rapidly in Phase I before beginning to stabilize in Phase II and becoming more stable Phase III. After two days of operation (and for the duration of the study), ammonia concentrations were below detection across all reactors by the end of the night.

Nutrient dynamics followed similar temporal trends in the three PBRs (Figure 3.1). All reactors exhibited undetectable NH_3-N by the end of the night after two days of operation (Table A.1) and maintained this performance for the remainder of the study. NO_2^- and NO_3^- were not detectable via ion chromatography when samples for the first 30 days of the experiment were tested. Initial biomass N:P was close to the Redfield ratio (7.26 g N·g^{-1} P) (Goldman et al., 1979) and did not differ between the three reactors (Phase I, p = 0.49). Although N:P curves differed during Phases II and III (p = 1.45E-11, p = 9.88E-7, respectively), they followed similar,

decreasing trends and stabilized in Phase III. Minimum N:P for each reactor ranged from 2.6 to 3.3 $g \cdot g^{-1}$, which approaches the proposed lower limit for cellular N:P (2.26 $g \cdot g^{-1}$) (Geider & LaRoche, 2002). These values are substantially lower than those reported for pure cultures of *Scenedesmus obliquus* (8.5) and *Chlamydomonas reinhardtii* (4.5) cultivated at an 8-day SRT with higher media N:P (7.84:1 vs. 2:1), demonstrating the inter- and intra-specific plasticity in biomass N:P (Gardner-Dale et al., 2017). Specific PO_4^{3-}-P uptake rates followed similar decreasing trends which stabilized in Phases II and III, but did not group together statistically (Phase I, FL and NC together, IL separate, p=0.00062; Phases II and III, p=6.74E-10, 5.283E-13, respectively). Culture density (VSS) continued to increase through time in all reactors. Final VSS values were 1,345 ± 41.59 $mg \cdot L^{-1}$, 1,466 ± 8 $mg \cdot L^{-1}$, and 1,679 ± 22.03 $mg \cdot L^{-1}$ for FL, IL, and NC, respectively. During Phases I and II, VSS differed in all three reactors ($p < 2.2E-16$) but began to converge in Phase III ($p = 0.000165$, IL and FL grouped together, NC separate). Effluent PO_4^{3-}-P and VSS data are presented in Figures A.3 and A.4, respectively. Overall performance of the three reactors was similar with respect to NH_3 and PO_4^{3-} dynamics, and was reflected in similar long-term trends in biomass N:P content.

3.3.2 Carbon Storage. Extant (i.e., observed during operation) carbohydrate and lipid storage exhibited similar temporal trends across reactors (Figure 3.2; A, B), with values which initially increased but subsequently decreased and stabilized with time. Carbohydrate content was initially similar in two of three reactors (Phase I; FL, IL together, NC separate, p = 3.06E-3), but fully diverged in Phase II ($p<2.2E-16$), before partially re-converging in Phase III (NC and FL together, IL separate, $p<2.2E-16$). Final carbohydrate content consistently increased from initial values (FL: 0.46 ± 0.015 to 0.98 ± 0.021, IL: 0.37 ± 0.008 to 0.77 ± 0.032, NC: 0.58 ± 0.026 g $carb \cdot g^{-1}$ protein), but was less than the maximum extant value for each reactor. Extant lipid content was initially similar for IL and NC, but not FL (Phase I, p = 6.06E-8), and diverged in Phases II and III ($p<2.2E-16$, p = 2.49E-5, respectively). While final lipid content in FL and IL had a net increase from SRT0 (FL: 0.32 ± 0.019 to 0.43 ± 0.047, IL: 0.23 ± 0.008 to 0.32 ± 0.006 026 g $lipid \cdot g^{-1}$ protein) to values less than their respective extant maxima, NC lipid content showed a net decrease with time (0.23 ± 0.008 to 0.22 ± 0.007 g $lipid \cdot g^{-1}$ protein). Carbohydrate and lipid content stabilized by the end of the experiment at which time carbohydrate and lipid values for the three reactors were within 32% and 42% of each other, respectively. When compared to the maximum intrinsic carbohydrate and lipid accumulation potentials measured during the kinetic experiment (Figure 3.2; C, D), this inter-reactor variation was small. Final extant carbohydrate content in the three reactors grouped together and were statistically

different from the intrinsic maximum (max = NC 144 hr, p<1E-8). SRT 8.75 lipid content of the three reactors differed both from each other and from the intrinsic maximum (max = FL 72hr, p<2.2E-16).

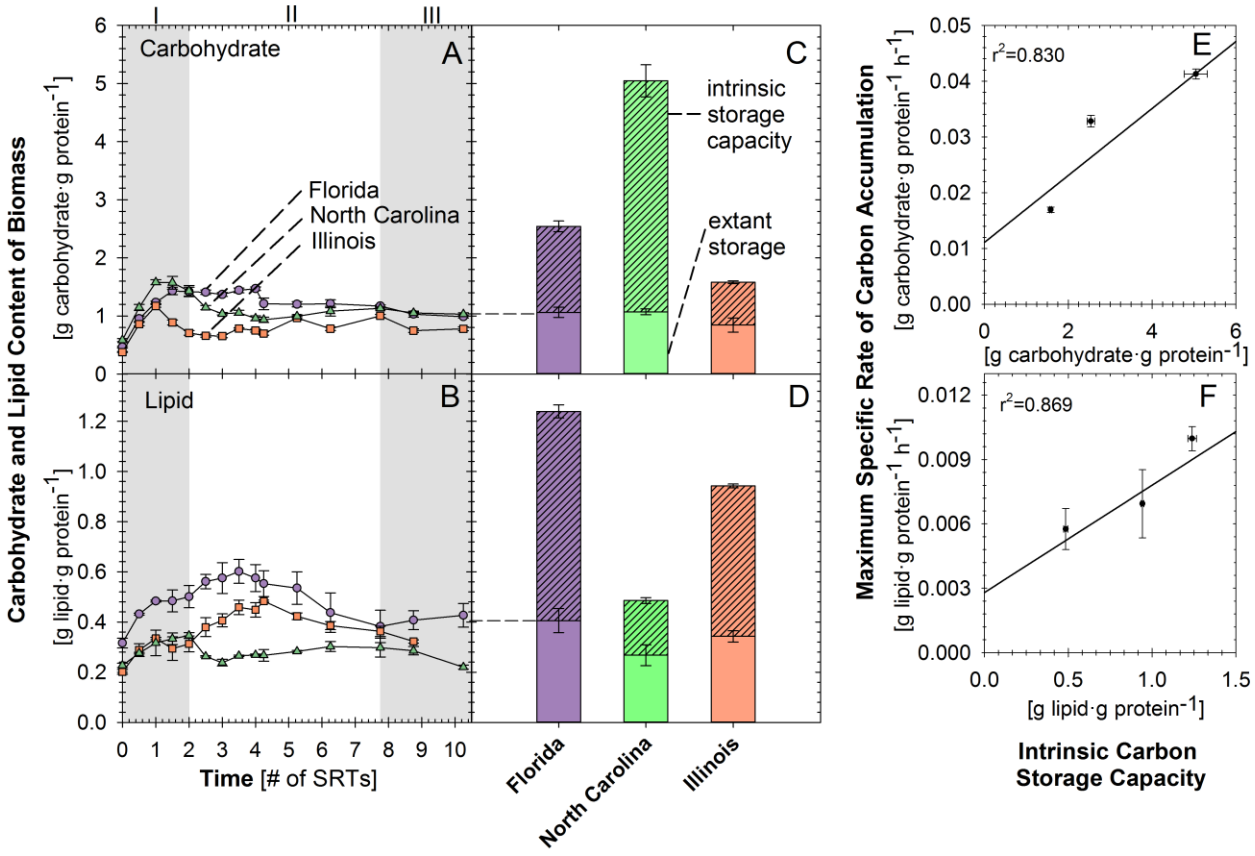

Figure 3.2: Extant (i.e., observed in the reactor) (A) carbohydrate and (B) lipid content of biomass through time in FL (purple circles), IL (orange squares), and NC (green triangles) reactors. Intrinsic (hatched) and extant (solid) (C) carbohydrate and (D) lipid storage, and the correlation between intrinsic carbohydrate (E) and lipid (F) storage capacity and maximum accumulation rate figure from the kinetic experiment. Extant carbohydrate and lipid storage stabilized in the three reactors by the end of the experiment. Variation in intrinsic storage capacity exceeded and was more variable than extant storage values. Maximum rate of carbohydrate or lipid accumulation and was positively correlated with intrinsic storage capacity of these storage products.

Table 3.1: Comparison of kinetic parameters from this study with literature values reported by Guest et al. (2013) and Mairet et al. (2011). With the exception of q_CH_max and f_LI_max values, which were lower, kinetic parameter values in this study were comparable to those reported in other studies.

Parameter	FL	IL	NC	Guest et al. (2013)	Mairet et al. (2011)	Units
Extant storage carbohydrate content ($f_{CH,extant}$)	0.17 ± 0.081	0.14 ± 0.072	0.10 ± 0.060	0.110	-	C-mol storage carbs · C-mol functional biomass $^{-1}$
Extant storage lipid content ($f_{LI,extant}$)	0.21 ± 0.0051	0.053 ± 0.014	0.027 ± 0.0017	0.09	-	C-mol storage lipid · C-mol functional biomass $^{-1}$
Maximum intrinsic storage carbohydrate accumulation capacity ($f_{CH,max}$)	0.76 ± 0.051	0.44 ± 0.067	1.48 ± 0.10	0.85	-	C-mol storage carbs · C-mol functional biomass $^{-1}$
Maximum intrinsic storage lipid accumulation capacity ($f_{LI,max}$)	0.86 ± 0.053	0.58 ± 0.070	0.18 ± 0.0087	1.1	-	C-mol storage carbs · C-mol functional biomass $^{-1}$
Maximum rate of storage carbohydrate accumulation ($q_{CH,max}$)	0.019 ± 0.00095	0.0087 ± 0.00037	0.020 ± 0.00041	0.031*	0.028**	*X_{CH}·mg(X_{CPO} as VSS)$^{-1}$·h^{-1} **mg carb·mg functional biomass^{-1}·h^{-1}
Maximum rate of storage lipid accumulation ($q_{LI,max}$)	0.0097 ± 0.00148	0.0087 ± 0.00061	0.0020 ± 0.00019	0.0091*	0.012**	*X_{LI}·mg(X_{CPO} as VSS)$^{-1}$·h^{-1} **mg lipid·mg functional biomass^{-1}·h^{-1}

In contrast to the noted similarities in nutrient dynamics and extant carbon accumulation, reactors differed widely in intrinsic maximum carbohydrate and lipid accumulation rate and capacity (Figure 3.2 C, D). To compare the results of this experiment with other values from the literature (Guest et al., 2013; Mairet et al., 2011), extant carbohydrate and lipid content, maximum rate of accumulation of storage carbohydrate and lipid, and maximum intrinsic functional carbohydrate and lipid storage capacity were recalculated as in Guest et al. (2013) (Table 3.1). Extant carbohydrate and lipid storage varied by less than two-fold and more than seven-fold, respectively (Table 1), but literature values fell within these observed ranges (Guest et al., 2013). Maximum rate of carbohydrate and lipid accumulation varied by two- and four-fold, respectively, between reactors. q_CH_max (Table 1) was 29% - 72% lower than literature

values (Guest et al., 2013; Mairet et al., 2011) but of the same order of magnitude (Table 1) while the observed range of q_LI_max spanned literature values (Table 1). Maximum carbohydrate and lipid storage potential varied more than three- and five-fold, respectively, between reactors. While f_CH_max values (Table 1) spanned that reported by Guest et al. (2013), f_LI_max was 22%-84%, and an order of magnitude lower than in Guest et al. (2013). This comparison suggests that daytime N-starvation in an SBR was not more effective, and in some cases was less effective than continually fed cyclostats maintained in P-limitation as in Guest et al. (2013) at inducing high extant or intrinsic storage capacity in mixed algal cultures.

Stronger positive correlations existed between kinetic parameters for lipids than for carbohydrates (Figure 3.2; E, F). Using values from Table 1, f_LI_extant correlated with q_LI_max and f_LI_max (r^2 = 0.955, 1.000, respectively), and q_LI_max with f_LI_max (r^2 = 0.959). There was no correlation between f_CH_extant and either q_CH_max or f_CH_max (r^2 = 0.082, 0.082, respectively), but f_CH_max and q_CH_max were somewhat correlated (r^2 = 0.621). This suggests that extant lipid content might be a predictor of lipid kinetic parameters, but it is unclear whether this relationship might hold for carbohydrates, given the similar values for f_CH_extant among the three reactors.

Reactors exhibited differences in intrinsic carbon storage rate and capacity after 10.25 SRTs, in contrast to nutrient dynamics and extant carbon storage which had more consistent values across reactors. These differences could be linked to differences in community structure which persisted after long-term operation under N-limited conditions. The following section presents the results from an 18S rRNA gene sequencing effort in which the taxonomic composition and culture dynamics of the reactors over the course of the long-term experiment are examined.

3.3.3 Community Analysis. Community richness estimates using V4 data were often approximately equal to or slightly less than V8-V9 values (Figure A.5). V8-V9 figures may be found in the SI (Figures A.6-A.7). Unless otherwise noted, V4 and V8-V9 results showed similar patterns. Community membership in the three reactors remained distinct throughout the experiment (Figure 3.3.A). AMOVA comparison using Bray Curtis distances and datasets divided corresponding to Phases I – III, yielded p-values ≤ 0.004 for both V4 and V8-V9 (Table A.2). AMOVA comparisons of Phases I-III within each reactor yielded p-values ≤0.001 for both V4 and V8-V9 in NC, but some higher values for FL and IL (V4:FL1-FL2 p=0.01, FL1-FL3 p=0.005; V8-V9: FL1-FL2 p=0.007, FL1-FL3 p=0.222, FL2-FL3 p=0.068, IL1-IL2, 0.004) (Table A.2).

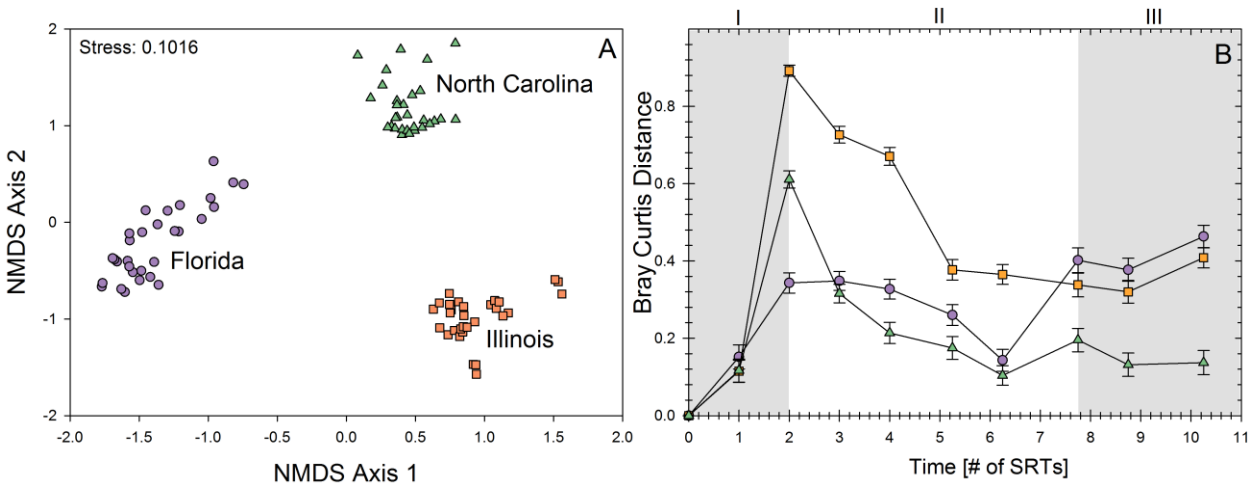

Figure 3.3: (A) NMDS plot based on Bray-Curtis dissimilarity matrix of all sample replicates and (B) Bray-Curtis distance from SRT0 in each reactor. Reactor communities remained distinct from one another throughout the experiment, while the community within each reactor diverged from its initial structure and then slowly returned to a community more similar to the starting community.

In all three reactors, the community rapidly diverged from SRT0 during Phase I, slowly returned to a community more closely resembling the starting community during Phase II, and stabilized during Phase III (with the exception of FL in the V8-V9 data, which shifted dramatically away again at SRT10.25). Indices of alpha diversity (nonparametric Shannon, Inverse Simpson, Chao1, Sobs) support this interpretation (Figure S8). FL and IL were similar in both number of observed species (Sobs) and Chao1 indices (Sobs = 16-35, Chao1=20-42), with NC ranges being much lower (Sobs=7-25, Chao1=7-30), corresponding with the fact that NC was inoculated only from surface water, while IL and FL also contained inoculum from WWTPs. Both the Inverse Simpson index and Nonparametric Shannon index were relatively stable with time for FL, but exhibited sharp transitions between Phases I and II and Phases II and III for IL and NC, indicating shared dominance between multiple OTUs in Phases I and III, but a community dominated by fewer OTUs during Phase II. These patterns are further supported by the relative read abundance data presented in Figure 3.4.

3.3.4 Taxonomic Analysis. Overall, between 88.2% and 99.7% of reads for each time point are included in the group comprised of the top 10 most abundant OTUs from each reactor (Figure 3.4). FL, NC, and IL had a total of 90, 56, and 88 OTUs each, which overlapped between reactors as follows: FL(59), IL(56), NC(33), FL-IL(12), IL-NC(4), FL-NC(3), FL-IL-NC(16). Thus, 13 of the 16 OTUs shared among all three reactors were in the top 10 most abundant for at least one reactor and the majority of OTUs contained a small cumulative fraction of reads and had little or no overlap between reactors. V4 OTU1, which classifies as *Acutodesmus*, was most abundant in all reactors and accounted for 45.8%, 68.2%, and 25.6% of overall reads in FL, NC, and IL, respectively (Figure 3.4). The most dominant V8-V9 OTU classified as *Coelastrella* (Figure A.7) which, along with *Acutodesmus*, is a member of the Scenedesmaceae. Other dominant OTUs represented a collection of green algae, diatoms, protozoa, amoeba, fungi, ciliates and rotifers (Table S4), indicating a multitrophic community (Arndt, 1993). For several OTUs, abrupt shifts in abundance occurred before SRT 2 and SRT 8, corresponding to Phases I-III described previously.

Copy number for the 18S rRNA gene varies from <10 to >10,000 (Vargas et al., 2015), causing difficulty in relating read counts with cell counts. Copy number is positively correlated on a log-log scale with cell diameter, although wide confidence intervals, particularly at larger cell sizes, prevent accurate interconversion of the two values (Vargas et al., 2015; Zhu et al., 2005). However, read counts may be interpreted as a proxy for cell biomass or biovolume, even among distantly related taxa (Vargas et al., 2015). Thus, the data in Figure 3.4 is best interpreted as representative of biomass distribution between OTUs within each reactor. Given that some of these OTUs represent taxa with larger cell sizes (eg. Rotifers, ciliates, protozoa) (Vargas et al., 2015), it seems likely that Figure 3.4 over-represents these taxa in terms of individual cells relative to smaller taxa such as green algae and diatoms (Vargas et al., 2015).

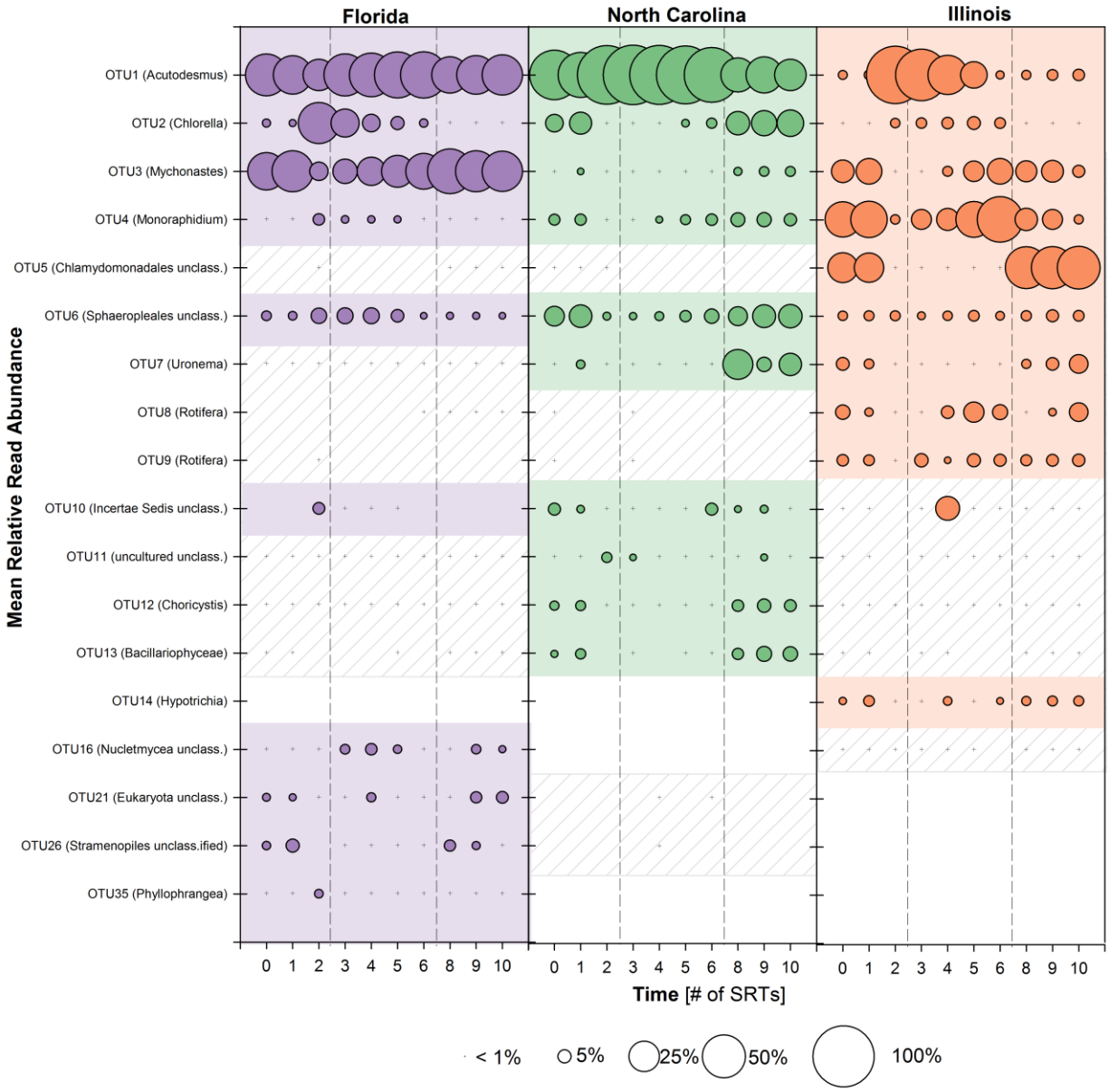

Figure 3.4: Mean relative read abundance of top 10 most abundant OTUs (colored shading) from each of the three reactors. Bubble area is proportional to mean relative abundance. Hatched shading indicates OTUs which were present but not abundant in a given reactor. Vertical dashed lines delineate Phases I, II, and III. SRTs are abbreviated as 1-10, but exact sample days are found in Table A.3 in the Appendix. The most abundant OTUs were shared between all three reactors and the majority of OTUs in the figure overlap between reactors, even if at low abundance. Shifts in relative abundance at SRT 2 and SRT 8 are visible for several OTUs, corresponding to transitions between Phases I-III as previously identified.

3.3.5 Comparison of Trends Across Datasets. All three reactors showed similar patterns of dynamic and stable periods with respect to many of the metrics presented in Figures 3.1-3.4. The first two SRTs (Phase I) were the time of greatest change in function as represented by nutrient dynamics (Figure 3.1) and carbohydrate content (Figure 3.2). The greatest shift in community structure (Figures 3.3.B, 3.4) occurred during the second SRT. PO_4^{3-}-P uptake and biomass composition (N:P, carbohydrate, lipid) remained dynamic for a subsequent period (Phase II) before stabilizing near the end of the experiment (Phase III). The community structure shifted back towards the starting community during Phase II (Figures 3.3.B, 3.4). Many OTUs strongly shifted in abundance between Phases II and III (Figure 3.4), but Bray Curtis distance from SRT0 increased only slightly in all reactors (Figure 3.3.B) and stabilized in Phase III. Due to similar patterns of change and stability observed over both structural and functional metrics, we sought to test for correlations between the two. However, a linear model relating taxonomic richness with functional metrics did not identify correlations which were consistently positive or negative across all three reactors using either V4 or V8-V9 data in either a combined model or when each functional metric was examined individually (Table A.4). This suggests that structure-function interactions are more complex than can be described by simple linear relationships. It is possible that the communities eventually reached a state of dynamic equilibrium characterized by stable function with continued structural dynamics, similar to those observed in activated sludge reactors where multitrophic microbial food webs drive unpredictable community dynamics while maintaining stable function (Kaewpipat & Grady, 2002; Kooi, Boer, & Kooijman, 1997). The research presented here supports the idea that algal communities of shifting composition may support stable function with respect to nutrient uptake and extant carbon accumulation.

3.4. Conclusions and Future Directions.

- Our results suggest that extant reactor performance which is temporally stable and similar between reactors may be achieved by algal communities which remain dynamic and distinct from one another through time.
- Daytime NH_3-N limitation might comprise a component of an engineered selective system for carbon accumulators, but numerous operational parameters must be co-optimized in order to maximize selection for carbohydrate or lipid-accumulating taxa, particularly in multitrophic systems.
- Systems designed to maximize storage product production by selecting for optimized intrinsic carbon storage capacity will require different design considerations than those

considered here, as reactors with similar and stable extant performance exhibited widely differing intrinsic performance.

**This concludes the research publication.*

Chapter 4: Conclusions and Engineering Significance

4.1 Nutrient Dynamics. The capacity for reliable nutrient removal is a key function of wastewater treatment system design, driven by both concern for environmental health and regulatory obligation. As has been described in previous sections, algae are well-suited to the task of N and P uptake, due to their flexible stoichiometry and potential to remove dissolved inorganic nutrients to the point of depletion. In this study, we demonstrated that three mixed algal cultures derived from geographically distinct locations and cultivated under identical laboratory conditions were capable of maintaining consistent inorganic N-limitation and biomass-specific PO_4^{3-}-P uptake for >10 SRTs. Integrated into a WWTP context, however, algae would be subject to variation in influent N:P ratios and nutrient speciation, which would be further influenced by the type and efficiency of other nutrient removal processes within the wastewater treatment system. Thus, it will be important to understand the factors dictating the behavior of algal systems under such varying conditions.

One challenge is to understand the interactions between SRT, algal community composition, and influent N:P and the combined impact of these factors on effluent quality. The influent N:P ratio used in this study (2:1 $g \cdot g^{-1}$ NH_3-N:PO_4^{3-}-P) was selected to ensure that the culture would be driven into N limitation. Biomass in this study was driven near the lower physiological limit for biomass N:P (Geider & LaRoche, 2002), while strains cultivated at higher media N:P (7.84:1), also at 8-day SRT, exhibited higher biomass N:P (approximately 4.5 and 9 for *C. reinhardtii* and *S. obliquus*, respectively) (Gardner-Dale et al., 2017). These results are generally in line with previous findings that biomass N:P will tend towards media N:P at low growth rates (Klausmeier et al., 2008). Achieving dual N and P limitation will require that influent N:P fall within the range of biomass nutrient plasticity for a given strain or community (Gardner-Dale et al., 2017). It will also be necessary to decipher the complex interactions between community structure, influent N:P, and SRT. In the likely scenario that influent nutrient fluxes and community composition would be dynamic on both short (hours-days) and long (seasonal) timescales, the design of processes aimed at achieving consistent dual limitation would be difficult. However, if at least one of these parameters could be reliably controlled, for example by pairing algal treatment with other nutrient removal methods or reliably constraining community composition, more reliable performance might be achieved.

4.2 Algae for Resource Recovery and Carbon Accumulation. Algal technologies have multiple attributes which fit nicely with novel concepts of WWTPs as ‘water resource recovery facilities’ (WRRFs) (Guest et al., 2009). Unlike in conventional WWTP technologies, N and P

assimilated into algal biomass remain bioavailable. Additionally, algal capacity for carbon accumulation as carbohydrate or lipid stores is reliably induced by nutrient limitation, making the production of carbohydrate- or lipid-rich algal biomass a natural choice of a co-process with algal wastewater treatment where nutrient removal is the primary goal.

Our operating system aimed to enhance the extant carbon-accumulating capacity of mixed cultures by applying the 'Survival of the Fattest' hypothesis proposed by Mooij et al. (2013) which reported success in enriching long-term extant carbohydrate storage in mixed cultures via daytime N-limitation and nighttime N-feeding. While our cultures exhibited rapid, short-term increases in both extant carbohydrate and lipid storage, these gains tapered off over time, with cultures exhibiting only moderate long-term increases in carbon storage which were far less than those reported by Mooij et al. (2016, 2013). Performance of the NC reactor in Phases I and II was most comparable to the results published by Mooij et al. (2016, 2013) in which high carbohydrate content (approaching 50%) and a single OTU were dominant. The NC reactor, like those operated by Mooij et al. but unlike the others in this study, was inoculated only from surface water. The use of MiSeq sequencing technology in this experiment, in contrast to PCR-DGGE by Mooij et al. (2013, 2015, 2016,), enabled the detection of numerous, lower-abundance OTUs. Additionally, slower growing taxa which would have washed out of the systems operated by Mooij et al. (SRT = 1.4 – 6 days) may have persisted under the longer 8-day SRT in this study. The community dynamics of the three reactors was also likely influenced by the startup period which, while intended to prevent algal washout due to low inoculum concentrations, may have initially selected for a starting community of fast-growing algae capable of rapid nutrient uptake and may have prevented some of the more dramatic long-term shifts in species composition observed by Mooij et al. (2016). While the three reactors did sustain moderate long-term gains in carbohydrate content and allow for stable nutrient recovery, application of the 'Survival of the Fattest' hypothesis is perhaps not the optimal means of maximizing carbohydrate-producing potential of a system.

A striking finding from this study, however, was that despite similar extant carbohydrate and lipid storage at the end of the long-term experiment, the three cultures varied widely in intrinsic storage potential and in their accumulation ratios of carbohydrate to lipid. The drastic increase between extant and intrinsic storage values implies that stressing cells in order to achieve maximum storage might be a desirable process to follow nutrient removal. A system could be designed to select for cells with high intrinsic, rather than extant, carbon storage, if these physiological characteristics could be made to confer increased fitness. However, since

maximum intrinsic storage capacity correlated positively with maximum storage rate for both carbohydrates and lipids, the latter might prove a useful metric for predicting maximum accumulation potential of given culture. Improved understanding of community structure-function relationships might also help to design systems with maximum storage capacity.

4.3 Persistent Challenges and Future Work. In order for algal treatment systems to transition from the lab to full-scale operation, there are many variables which must be accounted for. Wastewater treatment systems operate with continuous flow, unlike the SBRs used in this experiment. However, in another study, a mixed algal system operated with continuous flow maintained constant P-limitation and similar carbon-storage to the system run in this study (Guest et al., 2013), suggesting that the extreme SBR environment was not necessary to achieve consistent performance. Additionally, SRT and HRT in our system were equal. While an 8-day SRT is reasonable for algal systems, WWTP HRTs must be much shorter, often on the order of hours. While membrane technology is one way to decouple SRT and HRT, this option may be costly and clogging or fouling may present additional operational challenges.

Full-scale outdoor algal systems must be resilient to the effects of numerous additional short- and long-term fluctuations in environmental conditions which were not within the scope of our lab-scale study. While our study illustrated that mixed algal cultures derived from different latitudes could achieve consistent performance when subjected to identical laboratory cultivation conditions, an orthogonal study might investigate how the varying environmental conditions characteristic of different latitudes influence the performance of a common starting culture. Temperature, light intensity, and photoperiod would all vary on both short- and long-term time scales, the magnitude of which would vary based on geographic location. Thorough understanding of these factors will be important to the design of robust systems capable of consistent performance regardless of location.

An additional complicating factor is the presence of multiple trophic levels within mixed communities. While this study only examined the eukaryotic community, prokaryotes would undoubtedly play a role in the microbial food web and potentially in nutrient dynamics by fixing nitrogen in N-limited systems. While culture crash due to grazing did not occur in this study, it has been described as a concern by others (Kazamia et al., 2014). Multi-tiered trophic systems exhibit chaotic dynamics in activated sludge systems, but still support consistent reactor function (Kaewpipat & Grady, 2002; Kooi et al., 1997). It is possible that a similar phenomenon was observed in our reactors, in which fluctuating community dynamics supported stable nutrient uptake and extant carbon accumulation.

Improved capacity to link structural and functional metrics might still improve system design capabilities. In this study, combined and individual linear models were not able to successfully determine links between species richness and functional parameters. The elucidation of such relationships, if they exist, could provide valuable insights for system design. One factor which might confound this effort is that 18S rRNA gene copy number varies over four orders of magnitude among microbial eukaryotes (Vargas et al., 2015). While copy number scales roughly with cell size (Vargas et al., 2015; Zhu et al., 2005), it is difficult to make quantitative statements about community structure due to wide confidence intervals, making it currently impossible to accurately describe relative abundances of algal taxa using 18S sequencing data. One approach to solving this issue would be to identify another suitable gene with more stable copy number to use for sequencing of eukaryotic microbes, although the numerous benefits of the 18S gene were described in a previous section. Alternatively, a more exhaustive study of copy number variations between and within taxa should be undertaken in order to better model community structure from read count data.

4.4 Concluding Thoughts. Modern concerns about climate change, the depletion of natural resources, and environmental pollution make the adoption of paradigm shifts such as those represented by WRRFs more pressing than ever. Algal technologies are uniquely poised to address many of these concerns as recent developments in sequencing technology, modeling capability, and real-time data collection continue to expand our capacities for system design and control. However, as described above, numerous challenges remain. The combination of laboratory experiments such as the one presented here with process modeling and pilot-scale studies has the potential to streamline the development of algal biotechnology. After decades of research, the convergence of technological and scientific maturity, environmental necessity, and sociopolitical will may soon allow widespread implementation of algal systems integrating nutrient removal and biofuel production to become a reality.

References

Ågren, G. I. (2004). The C : N : P stoichiometry of autotrophs – theory and observations. *Ecology Letters*, *7*(3), 185–191. https://doi.org/10.1111/j.1461-0248.2004.00567.x

APHA, AWWA, WEF, Rice, E. W., Baird, R. B., Eaton, A. D., & Clesceri, L. S. (2012). *Standard Methods for the Examination of Water and Wastewater* (22nd ed.). Washington, DC: American Water Works Assn.

Arndt, H. (1993). Rotifers as predators on components of the microbial web (bacteria, heterotrophic flagellates, ciliates) — a review. *Hydrobiologia*, *255–256*(1), 231–246. https://doi.org/10.1007/BF00025844

Axelsson, M., & Gentili, F. (2014). A Single-Step Method for Rapid Extraction of Total Lipids from Green Microalgae. *PLoS ONE*, *9*(2), 1–6. https://doi.org/10.1371/journal.pone.0089643

Barthel, L., Oliveira, P. A. V. de, & Costa, R. H. R. da. (2008). Plankton biomass in secondary ponds treating piggery waste. *Brazilian Archives of Biology and Technology*, *51*(6), 1287–1298. https://doi.org/10.1590/S1516-89132008000600025

Bazin, P., Jouenne, F., Deton-Cabanillas, A.-F., Pérez-Ruzafa, Á., & Véron, B. (2014). Complex patterns in phytoplankton and microeukaryote diversity along the estuarine continuum. *Hydrobiologia*, *726*(1), 155–178. https://doi.org/10.1007/s10750-013-1761-9

Becker, E. W. (1994). *Microalgae Biotechnology and Microbiology*. Cambridge University Press.

Benita, Y., Oosting, R. S., Lok, M. C., Wise, M. J., & Humphery-Smith, I. (2003). Regionalized GC content of template DNA as a predictor of PCR success. *Nucleic Acids Research*, *31*(16), e99–e99. https://doi.org/10.1093/nar/gng101

Bott, C. B., & Parker, D. S. (2011). *Nutrient Management Volume II: Removal Technology Performance & Reliability. Water Environment Research Foundation, International Water Association.* Brown and Caldwell, Hampton Roads Sanitation District.

Bradley, I. M., Pinto, A. J., & Guest, J. S. (2016). Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. *Applied and Environmental Microbiology*, *82*(19), 5878–5891. https://doi.org/10.1128/AEM.01630-16

Brányiková, I., Maršálková, B., Doucha, J., Brányik, T., Bišová, K., Zachleder, V., & Vítová, M. (2011). Microalgae—novel highly efficient starch producers. *Biotechnology and Bioengineering*, *108*(4), 766–776. https://doi.org/10.1002/bit.23016

Cherkasov, N., Ibhadon, A. O., & Fitzpatrick, P. (2015). A review of the existing and alternative methods for greener nitrogen fixation. *Chemical Engineering and Processing: Process Intensification*, *90*, 24–33. https://doi.org/10.1016/j.cep.2015.02.004

Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnology Advances*, *25*(3), 294–306. https://doi.org/10.1016/j.biotechadv.2007.02.001

Clark, D. L., Hunt, G., Kasch, M. S., Lemonds, P. J., Moen, G. M., & Neethling, J. B. (2010). *Regulatory Approaches to Protect Water quality, Volume I - Review of Existing Practices.* Water Environment Research Foundation, International Water Association, HDR Engineering.

Collos, Y., & Harrison, P. J. (2014). Acclimation and toxicity of high ammonium concentrations to unicellular algae. *Marine Pollution Bulletin*, *80*(1), 8–23. https://doi.org/10.1016/j.marpolbul.2014.01.006

Conley, D. J., Paerl, H. W., Howarth, R. W., Boesch, D. F., Seitzinger, S. P., Havens, K. E., … Likens, G. E. (2009). Controlling Eutrophication: Nitrogen and Phosphorus. *Science*, *323*(5917), 1014–1015. https://doi.org/10.1126/science.1167755

Cordell, D., & White, S. (2014, October 17). Life's Bottleneck: Sustaining the World's Phosphorus for a Food Secure Future [review-article]. https://doi.org/10.1146/annurev-environ-010213-113300

Craggs, R. J., Davies-Colley, R. J., Tanner, C. C., & Sukias, J. P. (2003). Advanced pond system: performance with high rate ponds of different depths and areas. *Water Science and Technology*, *48*(2), 259–267.

Craggs, R. J., Heubeck, S., Lundquist, T. J., & Benemann, J. R. (2011). Algal biofuels from wastewater treatment high rate algal ponds. *Water Science and Technology*, *63*(4), 660–665. https://doi.org/10.2166/wst.2011.100

Craggs, Rupert J, Adey, W. H., Jessup, B. K., & Oswald, W. J. (1996). A controlled stream mesocosm for tertiary treatment of sewage. *Ecological Engineering*, *6*(1), 149–169. https://doi.org/10.1016/0925-8574(95)00056-9

Daigger, G. T. (2017). Flexibility and adaptability: essential elements of the WRRF of the future. *Water Practice and Technology*, *12*(1), 156–165. https://doi.org/10.2166/wpt.2017.019

Davis, R., Markham, J., Kinchin, C., Grundl, N., & Tan, E. C. D. (2016). Process Design and Economics for the Production of Algal Biomass: Algal Biomass Production in Open Pond Systems an drocessing Through Dewatering for Downstreamm Conversion. National Renewable Energy Laboratory, Golden CO. Retrieved from http://www.nrel.gov/docs/fy16osti/64772.pdf

Dragone, G., Fernandes, B. D., Abreu, A. P., Vicente, A. A., & Teixeira, J. A. (2011). Nutrient limitation as a strategy for increasing starch accumulation in microalgae. *Applied Energy*, *88*(10), 3331–3335. https://doi.org/10.1016/j.apenergy.2011.03.012

Eland, L. E., Davenport, R., & Mota, C. R. (2012). Evaluation of DNA extraction methods for freshwater eukaryotic microalgae. *Water Research*, *46*(16), 5355–5364. https://doi.org/10.1016/j.watres.2012.07.023

Folch, J., Lees, M., & Stanley, G. H. S. (1957). A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues. *Journal of Biological Chemistry*, *226*(1), 497–509.

Gardner-Dale, D. A., Bradley, I. M., & Guest, J. S. (2017). Influence of solids residence time and carbon storage on nitrogen and phosphorus recovery by microalgae across diel cycles. *Water Research*, *121*, 231–239. https://doi.org/10.1016/j.watres.2017.05.033

Geider, R. J., & LaRoche, J. (2002). Redfield Revisited: variability of C:N:P in marine microalgae and its biochemical basis. *European Journal of Phycology*, *37*(doi:10.1017/S0967026201003456), 1–17. https://doi.org/Geider, R.J. and LaRoche, Julie (2002) Redfield Revisited: variability of C:N:P in marine microalgae and its biochemical basis European Journal of Phycology, 37 . pp. 1-17. DOI 10.1017/S0967026201003456 <http://dx.doi.org/10.1017/S0967026201003456>.

Ghosh, A., Khanra, S., Mondal, M., Halder, G., Tiwari, O. N., Saini, S., … Gayen, K. (2016). Progress toward isolation of strains and genetically engineered strains of microalgae for production of biofuel and other value added chemicals: A review. *Energy Conversion and Management*, *113*, 104–118. https://doi.org/10.1016/j.enconman.2016.01.050

Ghosh, S., & Love, N. G. (2011). Application of rbcL based molecular diversity analysis to algae in wastewater treatment plants. *Bioresource Technology*, *102*(3), 3619–3622. https://doi.org/10.1016/j.biortech.2010.10.125

Godos, I. de, Blanco, S., García-Encina, P. A., Becares, E., & Muñoz, R. (2009). Long-term operation of high rate algal ponds for the bioremediation of piggery wastewaters at high loading rates. *Bioresource Technology*, *100*(19), 4332–4339. https://doi.org/10.1016/j.biortech.2009.04.016

Goldman, J. C., McCarthy, J. J., & Peavey, D. G. (1979). Growth rate influence on the chemical composition of phytoplankton in oceanic waters. *Nature*, *279*(5710), 210–215. https://doi.org/10.1038/279210a0

Gomaa, M. a., Al-Haj, L., & Abed, R. m. m. (2016). Metabolic engineering of Cyanobacteria and microalgae for enhanced production of biofuels and high-value products. *Journal of Applied Microbiology*, *121*(4), 919–931. https://doi.org/10.1111/jam.13232

González-Fernández, C., & Ballesteros, M. (2012). Linking microalgae and cyanobacteria culture conditions and key-enzymes for carbohydrate accumulation. *Biotechnology Advances*, *30*(6), 1655–1661. https://doi.org/10.1016/j.biotechadv.2012.07.003

Gu, A. Z., Nerenberg, R., Sturm, B. M., Chul, P., & Goel, R. (2010). Molecular Methods in Biological Systems. *Water Environment Research*, *82*(10), 908–930. https://doi.org/10.2175/106143010X12756668800735

Guest, J. S., Skerlos, S. J., Barnard, J. L., Beck, M. B., Daigger, G. T., Hilger, H., … Love*, N. G. (2009). A New Planning and Design Paradigm to Achieve Sustainable Resource Recovery from Wastewater. *Environmental Science & Technology*, *43*(16), 6126–6130. https://doi.org/10.1021/es9010515

Guest, J. S., van Loosdrecht, M. C. M., Skerlos, S. J., & Love, N. G. (2013). Lumped Pathway Metabolic Model of Organic Carbon Accumulation and Mobilization by the Alga Chlamydomonas reinhardtii. *Environmental Science & Technology*, *47*(7), 3258–3267. https://doi.org/10.1021/es304980y

Guo, L., Sui, Z., Zhang, S., Ren, Y., & Liu, Y. (2015). Comparison of potential diatom “barcode” genes (the 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta. *International Journal of Systematic and Evolutionary Microbiology*, *65*(Pt 4), 1369–1380. https://doi.org/10.1099/ijs.0.000076

Guschina, I. A., & Harwood, J. L. (2006). Lipids and lipid metabolism in eukaryotic algae. *Progress in Lipid Research*, *45*(2), 160–186. https://doi.org/10.1016/j.plipres.2006.01.001

Harwood, J. L., & Jones, A. L. (1989). Lipid Metabolism in Algae. *Advances in Botanical Research*, *16*, 1–53. https://doi.org/10.1016/S0065-2296(08)60238-4

Hendriks, A. T. W. M., & Langeveld, J. G. (2017). Rethinking Wastewater Treatment Plant Effluent Standards: Nutrient Reduction or Nutrient Control? *Environmental Science & Technology*. https://doi.org/10.1021/acs.est.7b01186

Hoffmann, J. P. (1998). Wastewater Treatment with Suspended and Nonsuspended Algae. *Journal of Phycology*, *34*(5), 757–763. https://doi.org/10.1046/j.1529-8817.1998.340757.x

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., & Darzins, A. (2008). Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *The Plant Journal: For Cell And Molecular Biology*, *54*(4), 621–639. https://doi.org/10.1111/j.1365-313X.2008.03492.x

Huber, J. A., Welch, D. B. M., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A., & Sogin, M. L. (2007). Microbial Population Structures in the Deep Marine Biosphere. *Science*, *318*(5847), 97–100. https://doi.org/10.1126/science.1146689

Kaewpipat, K., & Grady, C. P. L. (2002). Microbial population dynamics in laboratory-scale activated sludge reactors. *Water Science and Technology*, *46*(1–2), 19–27.

Kazamia, E., Riseley, A. S., Howe, C. J., & Smith, A. G. (2014). An Engineered Community Approach for Industrial Cultivation of Microalgae. *Industrial Biotechnology*, *10*(3), 184–190. https://doi.org/10.1089/ind.2013.0041

Klausmeier, C. A., Litchman, E., Daufresne, T., & Levin, S. A. (2008). Phytoplankton stoichiometry. *Ecological Research*, *23*(3), 479–485. https://doi.org/10.1007/s11284-008-0470-8

Kooi, B. W., Boer, M. P., & Kooijman, S. a. L. M. (1997). Complex dynamic behaviour of autonomous microbial food chains. *Journal of Mathematical Biology*, *36*(1), 24–40. https://doi.org/10.1007/s002850050088

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on

the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, *79*(17), 5112–5120. https://doi.org/10.1128/AEM.01043-13

Kunin, V., Engelbrektson, A., Ochman, H., & Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology*, *12*(1), 118–123. https://doi.org/10.1111/j.1462-2920.2009.02051.x

Li, Y., Leow, S., Fedders, A. C., Sharma, B. K., Guest, J. S., & Strathmann, T. J. (2017). Quantitative multiphase model for hydrothermal liquefaction of algal biomass. *Green Chemistry*, *19*(4), 1163–1174. https://doi.org/10.1039/C6GC03294J

Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, … Maggie Law. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine & Biotechnology*, *2012*, 1–11. https://doi.org/10.1155/2012/251364

Lohan, K. M. P., Fleischer, R. C., Carney, K. J., Holzer, K. K., & Ruiz, G. M. (2016). Amplicon-Based Pyrosequencing Reveals High Diversity of Protistan Parasites in Ships' Ballast Water: Implications for Biogeography and Infectious Diseases. *Microbial Ecology*, *71*(3), 530–542. https://doi.org/10.1007/s00248-015-0684-6

Lourenço, S. O., Barbarino, E., Marquez, U. M. L., & Aidar, E. (1998). Distribution of Intracellular Nitrogen in Marine Microalgae: Basis for the Calculation of Specific Nitrogen-to-Protein Conversion Factors. *Journal of Phycology*, *34*(5), 798–811. https://doi.org/10.1046/j.1529-8817.1998.340798.x

Mairet, F., Bernard, O., Masci, P., Lacour, T., & Sciandra, A. (2011). Modelling neutral lipid production by the microalga Isochrysis aff. galbana under nitrogen limitation. *Bioresource Technology*, *102*(1), 142–149. https://doi.org/10.1016/j.biortech.2010.06.138

Markou, G., Angelidaki, I., & Georgakakis, D. (2012). Microalgal carbohydrates: an overview of the factors influencing carbohydrates production, and of main bioconversion technologies for

production of biofuels. *Applied Microbiology & Biotechnology*, *96*(3), 631–645. https://doi.org/10.1007/s00253-012-4398-0

Medinger, R., Nolte, V., Pandey, R. V., Jost, S., Ottenwälder, B., Schlötterer, C., & Boenigk, J. (2010). Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular Ecology*, *19*, 32–40. https://doi.org/10.1111/j.1365-294X.2009.04478.x

Metcalf, & Eddy. (2014). *Wastewater Eineering Treatment and Resource Recovery* (5th ed.). New York, New York: McGraw-Hill Education.

Mohrbeck, I., Raupach, M. J., Martínez Arbizu, P., Knebelsberger, T., & Laakmann, S. (2015). High-Throughput Sequencing—The Key to Rapid Biodiversity Assessment of Marine Metazoa? *PLoS ONE*, *10*(10), 1–24. https://doi.org/10.1371/journal.pone.0140342

Mooij, P. R., Graaff, D. R. de, Loosdrecht, M. C. M. van, & Kleerebezem, R. (2015). Starch productivity in cyclically operated photobioreactors with marine microalgae—effect of ammonium addition regime and volume exchange ratio. *Journal of Applied Phycology*, *27*(3), 1121–1126. https://doi.org/10.1007/s10811-014-0430-3

Mooij, P. R., Jongh, L. D. de, Loosdrecht, M. C. M. van, & Kleerebezem, R. (2016). Influence of silicate on enrichment of highly productive microalgae from a mixed culture. *Journal of Applied Phycology*, *28*(3), 1453–1457. https://doi.org/10.1007/s10811-015-0678-2

Mooij, P. R., Stouten, G. R., Tamis, J., Loosdrecht, M. C. M. van, & Kleerebezem, R. (2013). Survival of the fattest. *Energy & Environmental Science*, *6*(12), 3404–3406. https://doi.org/10.1039/C3EE42912A

Mooij, P. R., Stouten, G. R., van Loosdrecht, M. C., & Kleerebezem, R. (2015). Ecology-based selective environments as solution to contamination in microalgal cultivation. *Current Opinion in Biotechnology*, *33*, 46–51. https://doi.org/10.1016/j.copbio.2014.11.001

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2017). vegan: Community Ecology Package. R Package version 2.4-3. Retrieved from https://CRAN.R-project.org/package=vegan

Oswald, W. J., Gotaas, H. B., Golueke, C. G., Kellen, W. R., Gloyna, E. F., & Hermann, E. R. (1957). Algae in Waste Treatment [with Discussion]. *Sewage and Industrial Wastes*, *29*(4), 437–457.

Paerl, H. W., Valdes, L. M., Joyner, A. R., Piehler, M. F., & Lebo, M. E. (2004). Solving Problems Resulting from Solutions: Evolution of a Dual Nutrient Management Strategy for the Eutrophying Neuse River Estuary, North Carolina. *Environmental Science & Technology*, *38*(11), 3068–3073. https://doi.org/10.1021/es0352350

Park, J. B. K., Craggs, R. J., & Shilton, A. N. (2011). Recycling algae to improve species control and harvest efficiency from a high rate algal pond. *Water Research*, *45*(20), 6637–6649. https://doi.org/10.1016/j.watres.2011.09.042

Park, J. B. K., Craggs, R. J., & Shilton, A. N. (2013). Enhancing biomass energy yield from pilot-scale high rate algal ponds with recycling. *Water Research*, *47*(13), 4422–4432. https://doi.org/10.1016/j.watres.2013.04.001

Pinto, A. J., Schroeder, J., Lunn, M., Sloan, W., & Raskin, L. (2014). Spatial-Temporal Survey and Occupancy-Abundance Modeling To Predict Bacterial Community Dynamics in the Drinking Water Microbiome. *mBio*, *5*(3), e01135-14. https://doi.org/10.1128/mBio.01135-14

Polz, M. F., & Cavanaugh, C. M. (1998). Bias in Template-to-Product Ratios in Multitemplate PCR. *Applied and Environmental Microbiology*, *64*(10), 3724–3730.

Pruvost, J., Van Vooren, G., Cogne, G., & Legrand, J. (2009). Investigation of biomass and lipids production with Neochloris oleoabundans in photobioreactor. *Bioresource Technology*, *100*(23), 5988–5995. https://doi.org/10.1016/j.biortech.2009.06.004

Rabalais, N. N., Turner, R. E., & Wiseman, W. J. (2002). Gulf of Mexico Hypoxia, a.k.a. "The Dead Zone." *Annual Review of Ecology and Systematics*, *33*, 235–263.

Redfield, A. C. (1958). The Biological Control of Chemical Factors in the Environment. *American Scientist*, *46*(3), 230A–221.

Reynolds, C. S., & Likens, G. E. (2009). Phytoplankton Population Dynamics: Concepts and Performance Measurement. In *Encyclopedia of Inland Waters*.

Rhee, G.-Y. (1978). Effects of N:P atomic ratios and nitrate limitation on algal growth, cell composition, and nitrate uptake 1. *Limnology and Oceanography*, *23*(1), 10–25. https://doi.org/10.4319/lo.1978.23.1.0010

Rittmann, B. E., & McCarty, P. L. (2001). *Environmental Biotechnology: Principles and Applications*. New York, New York: McGraw-Hill Higher Education.

Ronaghi, M. (2001). Pyrosequencing Sheds Light on DNA Sequencing. *Genome Research*, *11*(1), 3–11. https://doi.org/10.1101/gr.150601

Ruhl, H. A., & Rybicki, N. B. (2010). Long-term reductions in anthropogenic nutrients link to improvements in Chesapeake Bay habitat. *Proceedings of the National Academy of Sciences*, *107*(38), 16566–16570. https://doi.org/10.1073/pnas.1003590107

Schloss, P. D., & Westcott, S. L. (2011). Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. *Applied and Environmental Microbiology*, *77*(10), 3219–3226. https://doi.org/10.1128/AEM.02810-10

Shoener, B. D., Bradley, I. M., Cusick, R. D., & Guest, J. S. (2014). Energy positive domestic wastewater treatment: the roles of anaerobic and phototrophic technologies. *Environmental Science: Processes & Impacts*, *16*(6), 1204–1222. https://doi.org/10.1039/C3EM00711A

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic

community in marine anoxic water. *Molecular Ecology*, *19*, 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Sydney, E. B., da Silva, T. E., Tokarski, A., Novak, A. C., de Carvalho, J. C., Woiciecohwski, A. L., … Soccol, C. R. (2011). Screening of microalgae with potential for biodiesel production and nutrient removal from treated domestic sewage. *Applied Energy*, *88*(10), 3291–3294. https://doi.org/10.1016/j.apenergy.2010.11.024

Tenorio, R., Fedders, A. C., Strathmann, T. J., & Guest, J. S. (2017). Impact of Growth Phases on Photochemically Produced Reactive Species in the Extracellular Matrix of Algal Cultivation Systems. *Green Chemistry*.

US DOE. (2016). 2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy | Department of Energy. Retrieved July 7, 2017, from https://energy.gov/eere/bioenergy/downloads/2016-billion-ton-report-advancing-domestic-resources-thriving-bioeconomy

US EPA. (2002). Wastewater Technology Fact Sheet: Facultative Lagoons.

US EPA, OIG. (2008). Report: Despite Progress, EPA Needs to Improve Oversight of Wastewater Upgrades in the Chesapeake Bay Watershed [Reports and Assessments]. Retrieved May 18, 2017, from https://www.epa.gov/office-inspector-general/report-despite-progress-epa-needs-improve-oversight-wastewater-upgrades

US EPA, OW. (2011). Working in Partnership with States to Address Phosphorus and Nitrogen Pollution through Use of a Framework for State Nutrient Reductions [Overviews and Factsheets]. Retrieved May 18, 2017, from https://www.epa.gov/nutrient-policy-data/working-partnership-states-address-phosphorus-and-nitrogen-pollution-through

US EPA, OW. (2015). A Compilation of Cost Data Associated with the Impacts and Control of Nutrient Pollution [Reports and Assessments]. Retrieved May 18, 2017, from

https://www.epa.gov/nutrient-policy-data/compilation-cost-data-associated-impacts-and-control-nutrient-pollution

van Wychen, S., & Laurens, L. M. L. (2015). Determination of Total Carbohydrates in Algal Biomass: Laboratory Analytical Procedure (Technical Report No. NREL/TP-5100-60957). National Renewable Energy Laboratory, Golden CO.

Vanwonterghem, I., Jensen, P. D., Ho, D. P., Batstone, D. J., & Tyson, G. W. (2014). Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Current Opinion in Biotechnology*, *27*, 55–64. https://doi.org/10.1016/j.copbio.2013.11.004

Vargas, C. de, Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., … Karsenti, E. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605. https://doi.org/10.1126/science.1261605

Viprey, M., Guillou, L., Ferréol, M., & Vaulot, D. (2008). Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environmental Microbiology*, *10*(7), 1804–1822. https://doi.org/10.1111/j.1462-2920.2008.01602.x

Wen, C., Wu, L., Qin, Y., Nostrand, J. D. V., Ning, D., Sun, B., … Zhou, J. (2017). Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq platform. *PLOS ONE*, *12*(4), e0176716. https://doi.org/10.1371/journal.pone.0176716

Westcott, S. L., & Schloss, P. D. (2017). OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere*, *2*(2), e00073-17. https://doi.org/10.1128/mSphereDirect.00073-17

Zhu, F., Massana, R., Not, F., Marie, D., & Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, *52*(1), 79–92. https://doi.org/10.1016/j.femsec.2004.10.006

Appendix: Supplementary Materials for, "Impact of Daytime Nitrogen Limitation on Structure and Function of Mixed Microalgal Communities Derived from Three Geographically Distinct Sources"

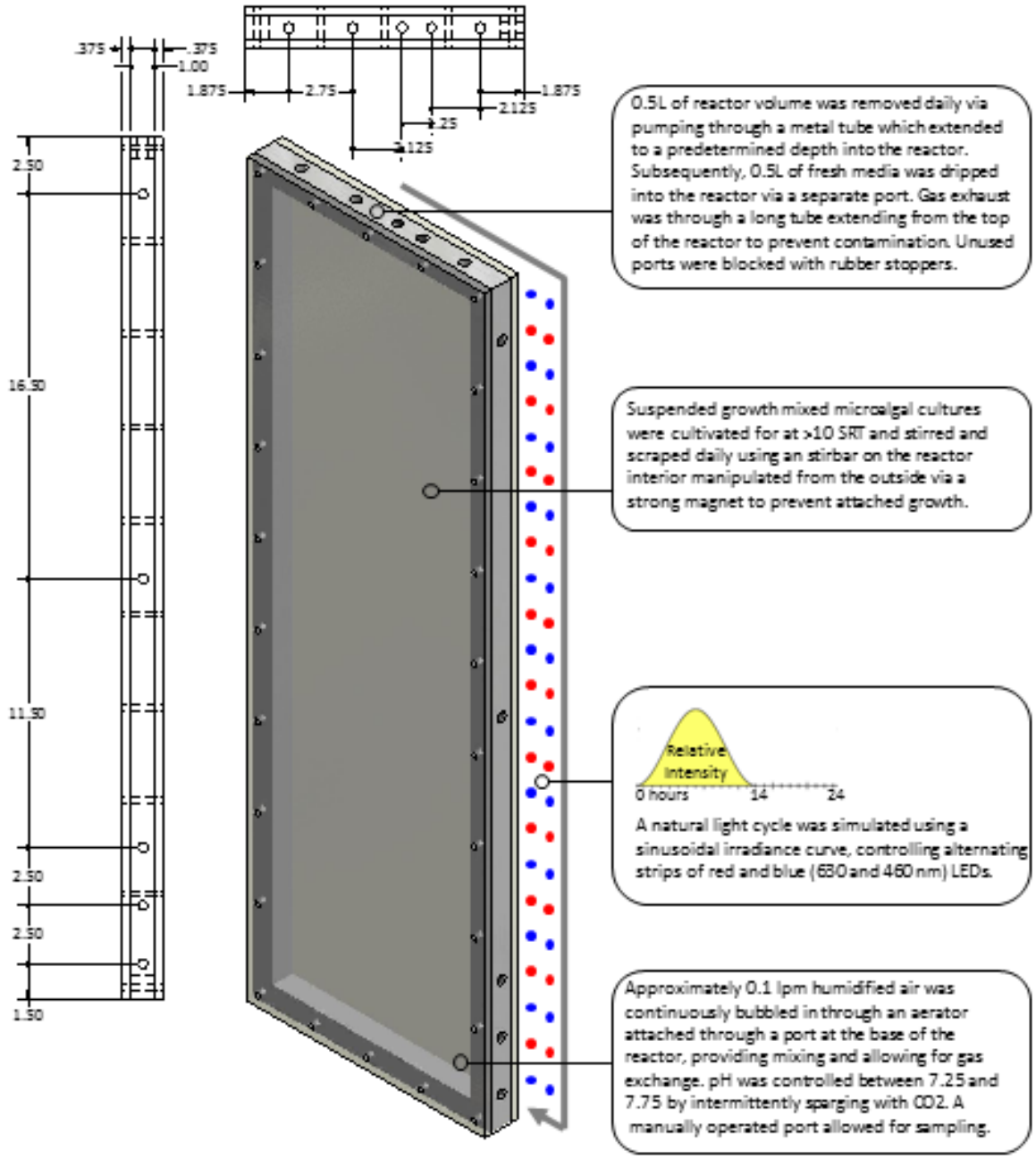

Figure A.1: Schematic illustrating photobioreactor and algal cultivation system design

A1. Reactor Inoculation and Startup Period. Inocula from each source (FL, NC, IL) were added to each of two PBRs (resulting in six PBRs in total during reactor startup), each containing 4L of modified TAP media(Tenorio, Fedders, Strathmann, & Guest, 2017) without TRIS buffer or acetic acid to minimize DOC, and including the following additions and modifications: $NaHCO_3$ (500 mg·L^{-1}), NH_4Cl (98.2 mg-N·L^{-1}), and PO_4^{3-} supplied as 2:1 (g·g^{-1}) K_2HPO_4:KH_2PO_4 (31.5 mg-P·L^{-1}), and Na_2SiO_3-$9H_2O$ (30 mg·L^{-1}). For acclimatization to laboratory lighting conditions, lighting intensity (at the surface of the reactors) was increased from 95 µE·m^{-2}·s^{-1} to 190 µE·m^{-2}·s^{-1} over the course of six days. One week after initial inoculation, the contents of each pair of reactors (two from FL, two from NC, two from IL) were mixed and re-divided. A sinusoidal light curve 14:10 (light:dark) was imposed with maximum light intensity of 190 µE·m^{-2}·s^{-1} and reactors were transitioned to sequencing batch mode with a hydraulic retention time (HRT) and solids residence time (SRT) of 8 days and cycle-time of 24 hours. Specifically, 0.5 L of reactor volume was wasted at the start of each dark phase (i.e., immediately after simulated sunset) and was immediately replaced with the addition of 0.5 L of new medium. One set of three reactors (one each of FL, NC, and IL) was operated with N-limited medium (58.9 mg-N·L^{-1} and 29.5 mg-P·L^{-1}; N:P mass ratio of 2:1) and the other with P-limited medium (186 mg-N·L^{-1} and 15.5 mg-P·L^{-1}; N:P mass ratio of 12:1). Media N:P ratios were set based on preliminary experiments with mixed communities (data not shown) and past experience with pure cultures.(Gardner-Dale et al., 2017) After seven days of operation in sequencing batch mode, the contents of each pair of reactors were again mixed and re-divided into two reactors, leaving other operating conditions unchanged. This represented time zero (t = 0 SRTs) for long-term operation.

A2. P-Limited Reactor Performance. P-limited reactors failed to reach P-limitation (data not shown), achieving minimum dissolved PO_4^{3-}-P between 0.10 and 3.14 ppm-P between experiment days 20 and 30 before increasing to between 12 and 14 mg PO_4^{3-}-P·L^{-1} by the end of the experiment. Similarly, reactor TSS steeply declined after 40 days of operation to 151-300 mg TSS·L^{-1}. Reactor washout and P-accumulation may have been due to NH_3 toxicity. NH_3 reached a maximum near 100 ppm NH_3-N on day 20, and maintained concentrations between approximately 80 and 100 ppm NH_3-N for the remainder of the experiment. 100 ppm NH_3-N (7.14 mM NH_3) exceeds concentrations known to be toxic to numerous algal taxa (e.g., Dinophyceae, Diatomophyceae, Raphidophyceae, and Prymnesiophyceae) and inhibitory to others (eg. Cyanophyceae) (Collos & Harrison, 2014). While some taxa, such as Chlorophyceae, are reported to tolerate NH_3 levels up to 39 mM (Collos & Harrison, 2014), it seems probable that the combination of ammonia toxicity and stress significantly impaired

system function. This might be remedied in future studies by reducing the media NH_3:PO_4^{3-} media, or using an alternative N source such as NO_3^-. Only data collected from the N-limited reactors are presented here.

A3. Mock Community Analysis. Read overlap and clustering threshold for V4-sequenced mock community samples corresponding to mock community M4 (Bradley et al., 2016) were varied between 10 and 90 base pairs (bp) and 0-9%, respectively. Sequence data was compared to a theoretical reference sample consisting of two sequences of each of the twelve MC members. The Jaccard dissimilarity index between each sequenced sample and reference sample at each overlap and clustering threshold was used to create a cost-benefit matrix (Figure A.2). The minimum clustering threshold and overlap combination yielding a Jaccard dissimilarity index of ≤ 0.2 was 5% clustering and an overlap of 70 bp. Sequencing error was calculated to be 0.00588% using the seq.error command in Mothur. When these parameters were used to analyze the bulk data, it was not possible to resolve closely related taxa which are commonly found in wastewater (eg. *Acutodesmus* and *Monoraphidium*). Since 3% clustering is commonly used in the literature, we used 3% clustering and 70 bp overlap for the final data analysis. Since community M4 was designed to encompass the wide diversity of freshwater and marine eukaryotic microalgae, these results highlight the need in future work to select mock community members based on the expected algal community to ensure adequate resolution.

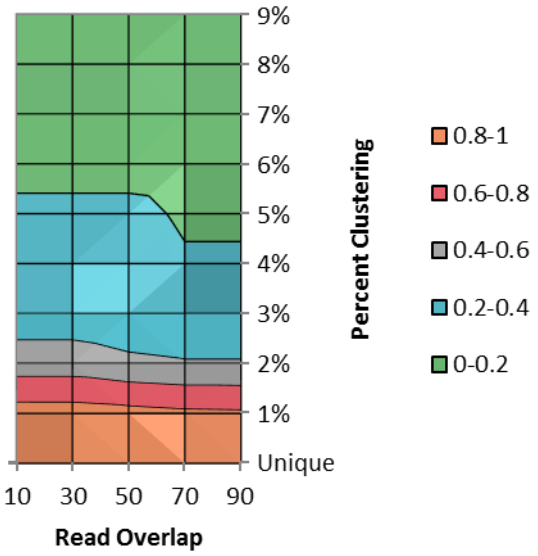

Figure A.2: Cost-benefit analysis of read overlap (x-axis) and percent clustering (y-axis). Colors indicate values of Jaccard dissimilarity index between sequenced mock community and a theoretical sample containing two correct sequences for each mock community member.

Table A.1: With the exception of the samples in this table, all reactors had undetectable NH_3-N by the end of the night for the entire long-term experiment.

Reactor	Experiment Day	End of Night NH_3-N (ppm)	Standard Deviation
IL	0	7.55	0.021
FL	0	0.59	0.034
IL	1	5.62	0.092

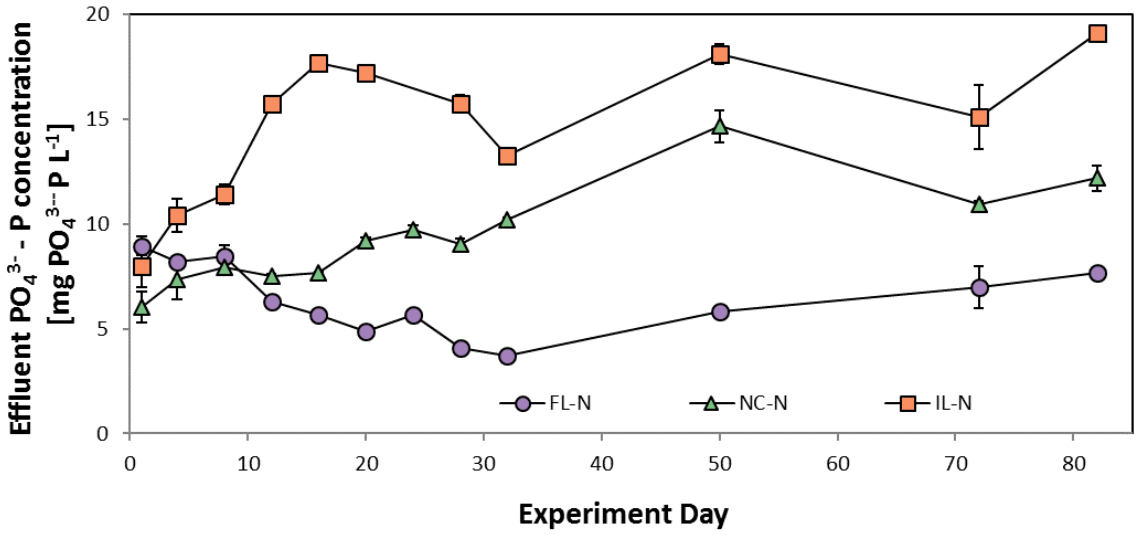

Figure A.3: PO_4^{3-}-P concentration of reactor effluent sampled at end of light period.

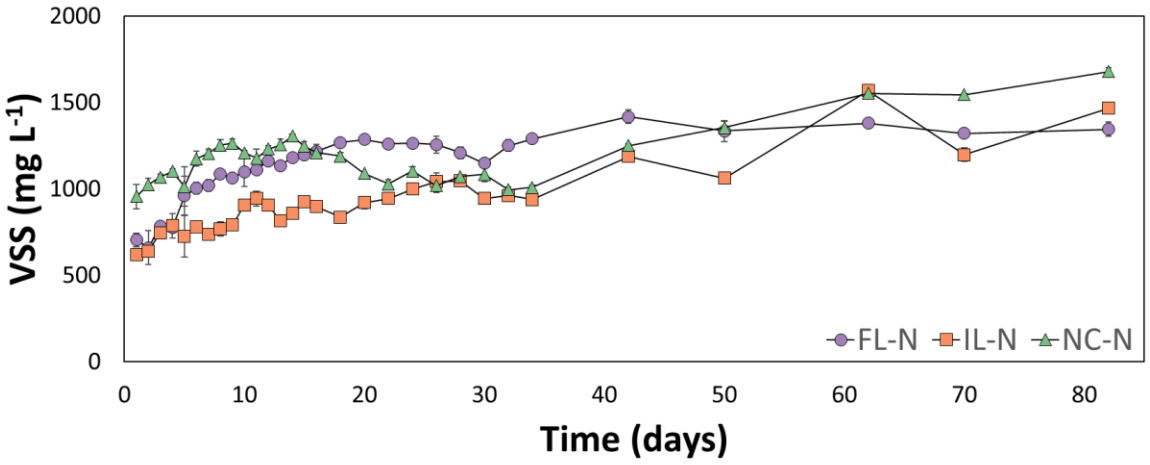

Figure A.4: Effluent volatile suspended solids (VSS) measured at the end of the light period.

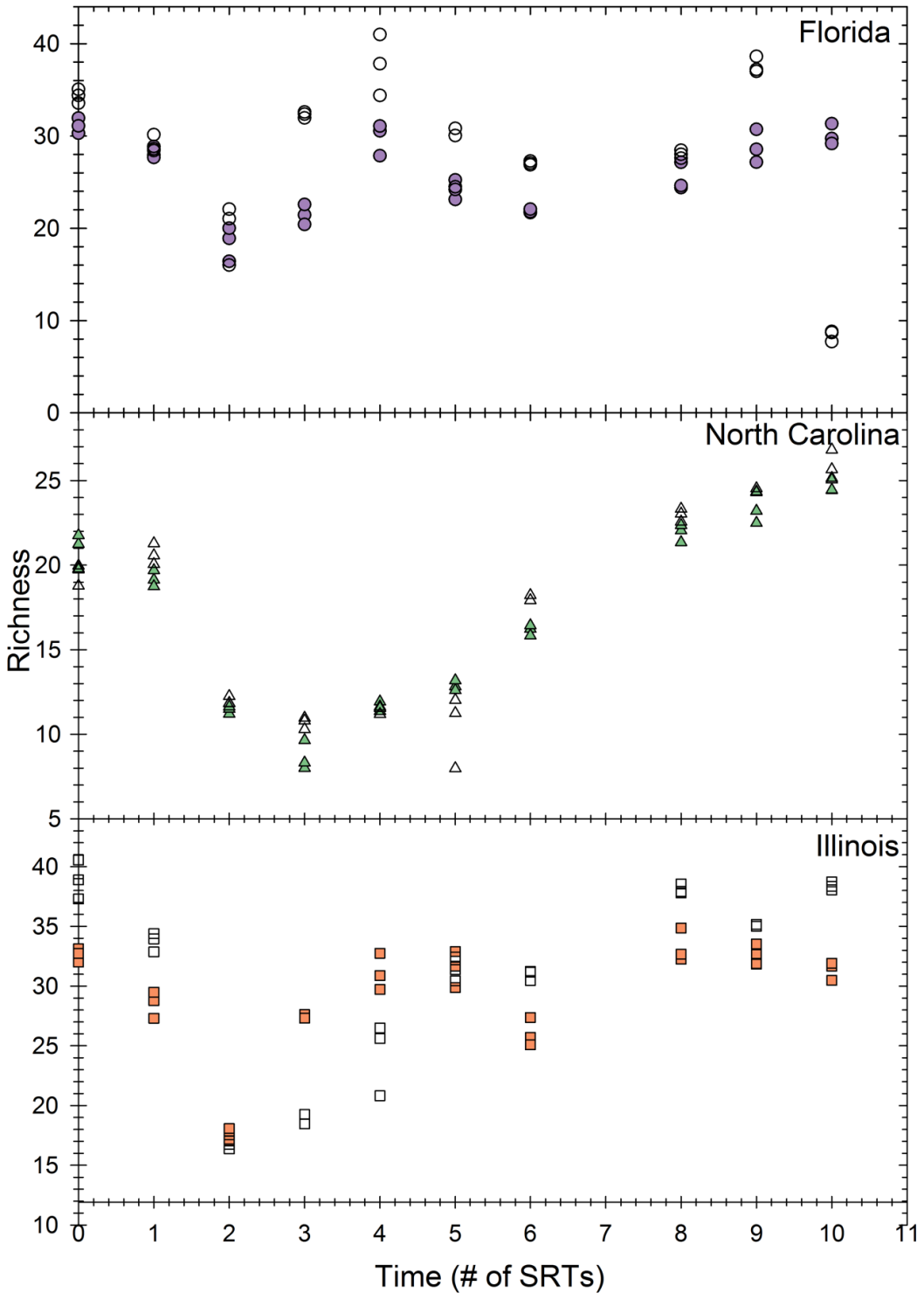

Figure A.5: Comparison of expected community richness of an OTU table rarefied to minimum sample size between V4 (filled symbols) and V8-V9 (empty symbols) through time in each of the three reactors shows that V4 sequencing most often yielded richness measurements which were approximately equal to or slightly lower than V8-V9 values.

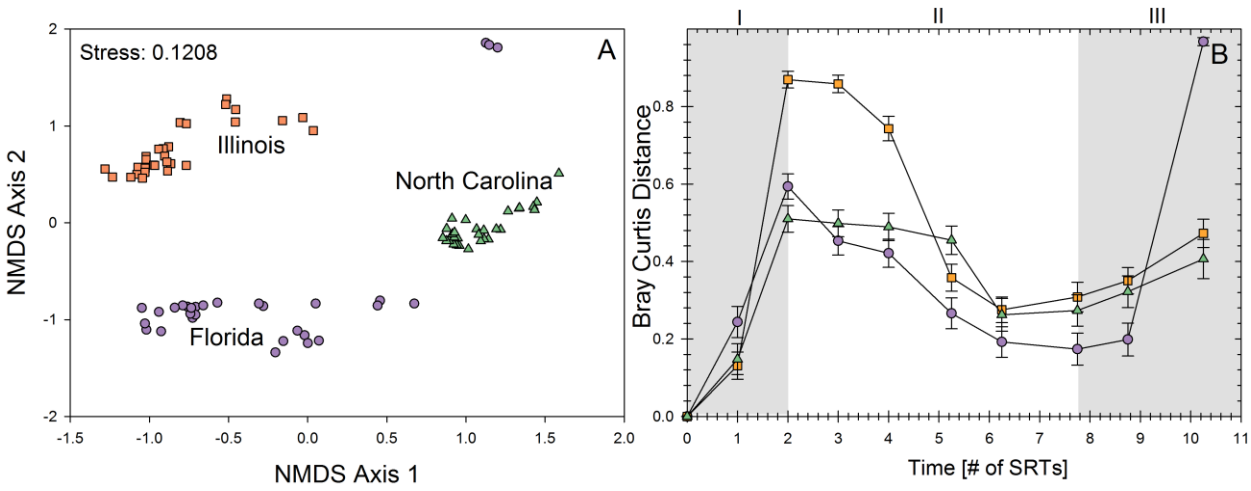

Figure A.6: NMDS Plot (A) of V8-V9 Bray-Curtis dissimilarity between all sequenced samples and (B) Bray-Curtis distance from SRT0 in each reactor through time. With the exception of the last point on FL SRT10, all reactors followed similar trends of increased distance from SRT0 followed by slow return to a community more closely resembling the initial community.

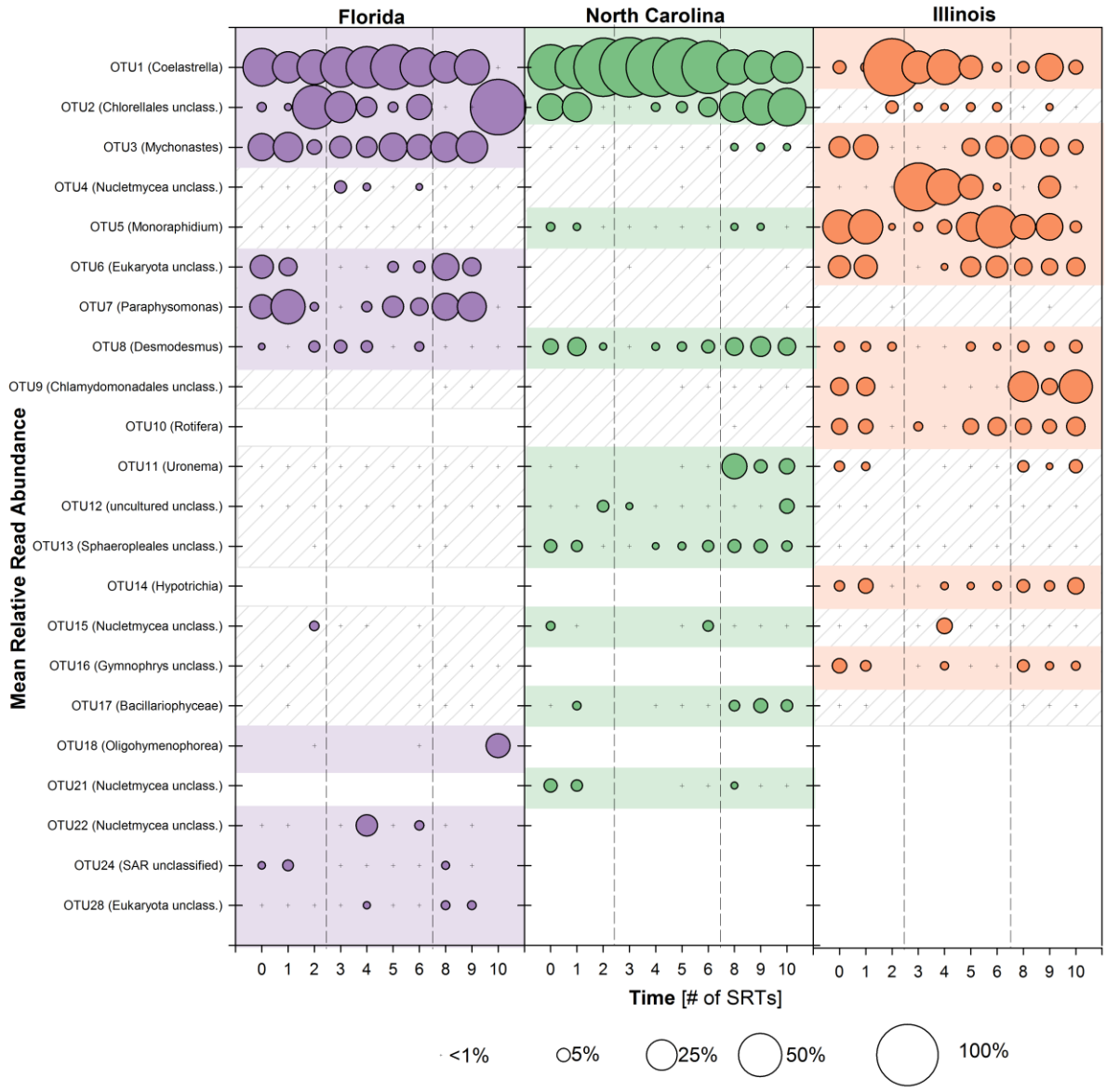

Figure A.7: Relative abundance of top 10 most abundant OTUs (colored shading) identified by V8-V9 sequencing in each of the three reactors. Bubble area is proportional to mean relative abundance. Hatched shading indicates taxa which were present but not abundant in a given reactor. V8-V9 OTU1, which classifies as Coelastrella, represented 35.4%, 65.3%, and 22.0% of reads in FL, NC, and IL, respectively. Between 82.5% and 99.6% of reads for each sample point were accounted for in the 22 OTUs represented in this figure. FL, NC, and IL contained a total of 126, 54, and 100 OTUs each, respectively, with OTUs overlapping between reactors as follows: FL(85), IL(57), NC(26), FL-IL(17), IL-NC(4), FL-NC(2), FL-IL-NC(22). 14 of the 22 OTUs which were shared among all reactors are represented in Figure S7.

Table A.2: V4 and V8-V9 AMOVA results using Bray Curtis distances to compare Phases I, II, and III within and between the three reactors. 'Comparison' column indicates the reactors and phases being compared (eg. FL_1 corresponds to FL, Phase I).

Comparison	V4 p-value	V8-V9 p-value
F1-F2	0.01	0.007
F1-F3	0.005	0.222
F2-F3	<0.001	0.068
I1-I2	0.001	0.004
I1-I3	<0.001	<0.001
I2-I3	<0.001	<0.001
N1-N2	<0.001	<0.001
N1-N3	<0.001	0.001
N2-N3	<0.001	<0.001
F1-I1	0.004	0.002
F1-N1	0.001	0.001
I1-N1	0.001	0.002
F2-I2	<0.001	<0.001
F2-N2	<0.001	<0.001
I2-N2	<0.001	<0.001
F3-I3	<0.001	<0.001
F3-N3	<0.001	0.004
I3-N3	<0.001	0.001

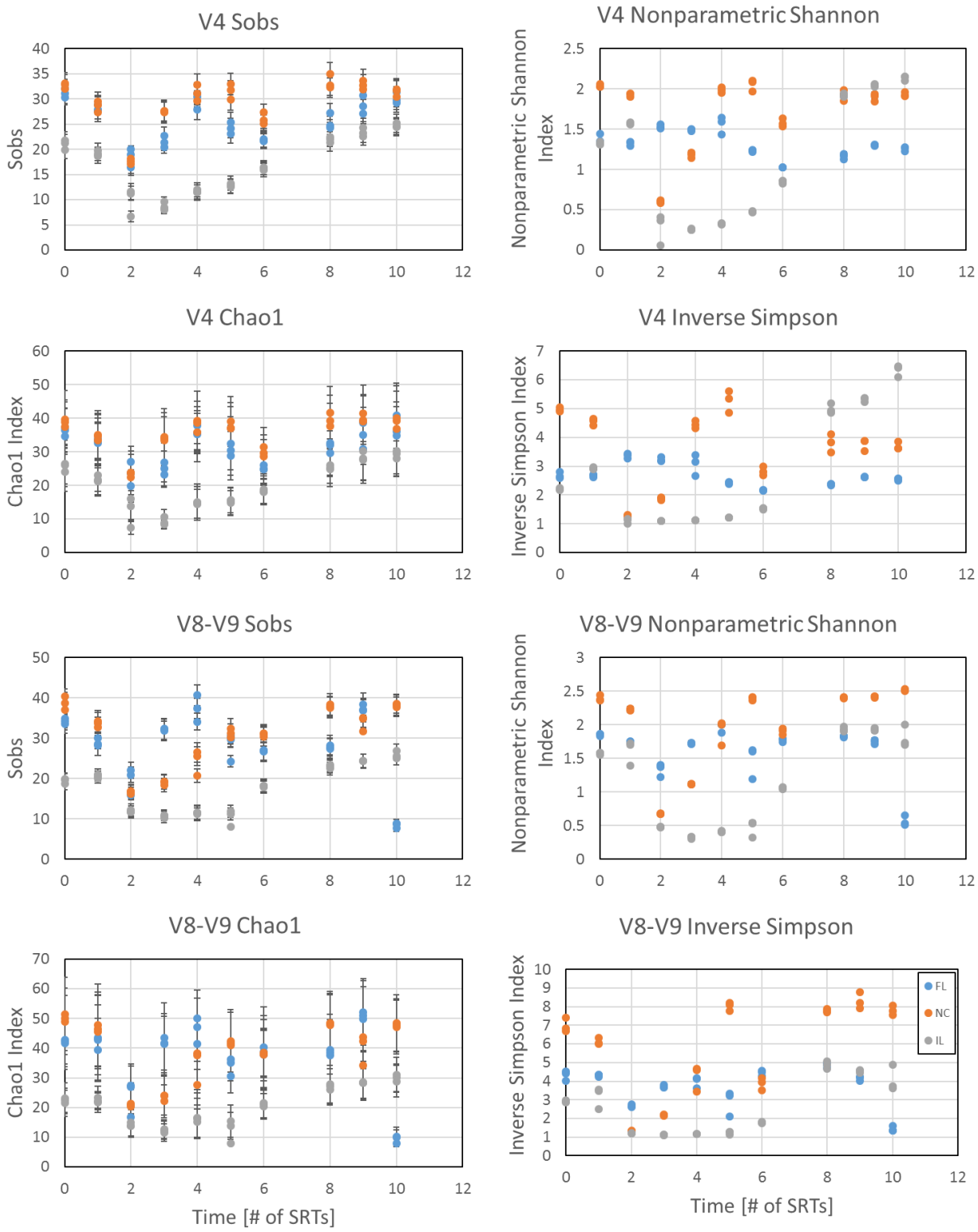

Figure A.8: Alpha diversity metrics including Chao1, Sobs, inverse Simpson, and nonparametric Shannon indices for the three reactors over the course of the long-term experiment using V4 and V8-V9 sequencing data.

Table A.3: Experiment days corresponding to x-axis OTU designations for main text Figure 3.4

SRT label	Experiment Day
0	0
1	8
2	16
3	24
4	32
5	40
7	54
8	62
9	70
10	82

Table A.4: Taxonomic classifications for most abundant 10 OTUs as determined by V4 (Figure 3.4) and V8-V9 (Figure A.7) sequencing analysis.

Primer Set	V4 Group	Taxon	Classification
V4	Otu001	Acutodesmus	Algae
V4	Otu002	Chlorella	Algae
V4	Otu003	Mychonastes	Algae
V4	Otu004	Monoraphidium	Algae
V4	Otu005	Chlamydomonadales_unclassified	Algae
V4	Otu006	Sphaeropleales_unclassified	Algae
V4	Otu007	Uronema	Algae
V4	Otu008	Rotifera	rotifer
V4	Otu009	Rotifera	rotifer
V4	Otu010	Incertae_Sedis_unclassified	unknown
V4	Otu011	uncultured_unclassified	unknown
V4	Otu012	Choricystis	Algae
V4	Otu013	Bacillariophyceae	diatom
V4	Otu014	Hypotrichia	protozoa
V4	Otu016	Nucletmycea_unclassified	fungi
V4	Otu021	Eukaryota_unclassified	unknown
V4	Otu026	Stramenopiles_unclassified	algae or diatoms
V4	Otu035	Phyllophrangea	ciliate
V8-V9	Otu001	Coelastrella	algae
V8-V9	Otu002	Chlorellales unclass.	algae
V8-V9	Otu003	Mychonastes	algae
V8-V9	Otu004	Nucletmycea unclass.	fungi
V8-V9	Otu005	Monoraphidium	algae
V8-V9	Otu006	Eukaryota unclass.	unknown
V8-V9	Otu007	Paraphysomonas	flagellate
V8-V9	Otu008	Desmodesmus	algae
V8-V9	Otu009	Chlamydomonadales unclass.	algae
V8-V9	Otu010	Rotifera	rotifer
V8-V9	Otu011	Uronema	algae
V8-V9	Otu012	uncultured unclass.	unknown
V8-V9	Otu013	Sphaeropleales unclass.	algae
V8-V9	Otu014	Hypotrichia	protozoa
V8-V9	Otu015	Nucletmycea unclass.	fungi
V8-V9	Otu016	Gymnophrys unclass.	protist
V8-V9	Otu017	Bacillariophyceae	diatom
V8-V9	Otu018	Oligohymenophorea	ciliate
V8-V9	Otu021	Nucletmycea unclass.	fungi
V8-V9	Otu022	Nucletmycea unclass.	fungi
V8-V9	Otu024	SAR unclassified	unknown
V8-V9	Otu028	Eukaryota unclass.	unknown

Table A.5: Results from a combined linear model (A, B) incorporating species richness and functional metrics and of correlations (C, D) between species richness and each functional metric individually.

A. V4 Combined Linear Model

Functional Parameter	FL_slope	p-value	IL_slope	p-value	NC_slope	p-value
Biomass N:P	-1.40E+01	0.2922	769.19	4.07E-06	5.95E+00	3.70E-05
Biomass Carbohydrate:Protein	-3.94E+01	0.0308	-32194.3	3.39E-06	-2.51E+00	0.34128
Biomamss Lipid:Protein	3.69E+01	0.3105	15991.29	3.33E-06	4.38E+01	0.00299
Biomass Normalized PO4-P uptake	-1.11E+03	0.1799	-288937	3.51E-06	8.38E+02	0.00379
Volatile Suspended Solids	-9.47E-02	0.2389	NA	NA	5.57E-02	6.03E-08
Media Dissolved Organic Carbon	-5.68E-02	0.6421	NA	NA	2.06E-01	0.00382

B. V8-V9 Combined Linear Model

Functional Parameter	FL_slope	p-value	IL_slope	p-value	NC_slope	p-value
Biomass N:P	58.80568	0.001275	-539.3	0.00407	3.621317	0.000964
Biomass Carbohydrate:Protein	34.55246	0.098721	21352.9	0.00503	-0.53713	0.80733
Biomamss Lipid:Protein	190.0284	0.000282	-10602.1	0.00499	4.977959	0.647566
Biomass Normalized PO4-P uptake	3534.657	0.0015	193569.7	0.00485	569.8045	0.015272
Volatile Suspended Solids	0.35662	0.001195	NA	NA	0.039865	5.93E-07
Media Dissolved Organic Carbon	1.04609	1.10E-06	NA	NA	0.288892	3.43E-05

C. V4 Individual Correlations

Functional Parameter	FL_slope	p-value	IL_slope	p-value	NC_slope	p-value
Biomass Carbohydrate:Protein	0.056	0.001	0.01	> 0.05	-0.14	> 0.05
Biomass N:P	0.45	0.01	-0.31	0.05	0.1	> 0.05
Media Dissolved Organic Carbon	-0.28	> 0.05	-0.51	0.001	0.47	0.001
Biomass Normalized PO4-P uptake	0.35	0.05	0.12	> 0.05	-0.42	0.01
Biomamss Lipid:Protein	-0.4	0.01	-0.02	> 0.05	-0.19	> 0.05
Volatile Suspended Solids	-0.37	0.01	0.34	0.05	0.64	0

D. V89 Individual Correlations

Functional Parameter	FL_slope	p-value	IL_slope	p-value	NC_slope	p-value
Biomass Carbohydrate:Protein	-0.02	> 0.05	0.14	> 0.05	-0.01	> 0.05
Biomass N:P	0.19	> 0.05	-0.43	0.01	-0.05	> 0.05
Media Dissolved Organic Carbon	0.46	0.01	-0.72	0	0.62	0
Biomass Normalized PO4-P uptake	0.33	> 0.05	0.19	> 0.05	-0.59	0.001
Biomamss Lipid:Protein	0.08	> 0.05	-0.4	0.01	-0.17	> 0.05
Volatile Suspended Solids	-0.26	> 0.05	0.3	> 0.05	0.72	0